From Idea to Product - Translating Knowledge between the Lab and the Clinic

FROM IDEA TO PRODUCT – TRANSLATING KNOWLEDGE BETWEEN THE LAB AND THE CLINIC

Robert S. Huckman **Ayfer H. Ali**

ABSTRACT

This dissertation is composed of three essays looking at innovation at Academic Medical Centers. It tries to empirically explore the problem of translating knowledge from the laboratory bench to the clinic and from the clinic to the bench.

Chapter 1, co-authored with Iain Cockburn, establishes the importance of in-house complementary knowledge in firm decision to license an invention from an Academic Medical Center. By using patent data to describe the technology portfolio of firms who look at patents and then decide whether to license them or not we are able to provide a description of demand in Markets for Technology. We show that firms license inventions that are similar to own technology portfolio when such similarity is measured at a broad level using International Patent Classes. However, controlling for such broad level proximity, firms are less likely to license inventions that are similar when measured at a more granular level.

Chapter 2 asks: “Are inventions by teams from Academic Medical Centers that combine cross-domain knowledge at a higher hazard of licensing than inventions by single domain teams?” Inventors’ educational background is used to assign them to the clinical (MDs) or the research domain (PhD). Contrary to our expectations, we find that inventions by cross-domain teams are at a lower hazard of licensing. Similarly, inventions by cross-domain integrated teams (at least one MD/PhD) are at a lower hazard of licensing than inventions by cross-domain distributed teams (MD and PhD on team but no MD/PhD). However, medical device inventions tend to be at a higher hazard of licensing if invented by cross-domain teams.

Chapter 3, co-authored with Rob Huckman, looks at how the routine clinical work of cardiac surgeons at Academic Medical Centers can impact their innovative performance as measured by quantity and quality of academic articles that they publish. We use the procedures that these cardiac surgeons perform every year to create a measure of clinical focus to understand whether diversity of work impacts innovation. Using a panel data with surgeon fixed effects we find that early career surgeons benefit from work diversity but late-career surgeons do not.

CONTENTS

Dedicated to the Memory of

my mother Zeynep Musova Alieva, who keeps inspiring me

and

my grandfather Musa Aliev, who valued education as much as he valued family

ACKNOWLEDGEMENTS

I could not have completed this dissertation without my mentor Iain Cockburn whose encouragement and dedication have made my last two years more productive and joyful. I would also like to thank my dissertation committee members – Rob Huckman, Carliss Baldwin and David Cutler. A debt of gratitude also goes to Dan Johnson, the person who taught me to love research.

While working on this thesis I have received inspiration, encouragement and insight from many special people: Harold Demonaco, Eric von Hippel, Karim Lakhani, Michelle Gittelman, Tim Simcoe, Jeff Furman, Brent Goldfarb, Scott Stern, Ajay Agarwal, Frances Toneguzzo, Seema Basu and Rachelle Safran.

My amazing friends Patricia Pepper and Jean Adams have been there for me in the difficult times and made me wiser, more spiritual and more fun.

Finally, I thank my unusual and amazing family: Fatme Ali, Musa Ali, Dzhemile Ali, Seher Yozgyur, Zeynep Yozgyur and Ilhan Yozgyur.

1. BUYER BEHAVIOR IN MARKETS FOR TECHNOLOGY: TECHNOLOGY PROXIMITY BETWEEN FIRM PORTFOLIO AND IN-LICENSED PATENTS

Ayfer Ali **Iain Cockburn**

1.1 Abstract

Markets for technology promise to increase productivity by better allocating innovative capacity across firms. Research on the demand side of these markets, however, has been limited. In this paper, we use a new dataset of patents available for licensing from a large, innovative academic medical center (AMC) to understand the structure of these markets. Our data includes information on all firms that showed interest in these patents by signing a confidentiality agreement and later decided whether to license or not license the focal technology. Strikingly, we find that of the 285 patents we observe, about 30% of patents available for licensing are never even looked at, and of those that are looked at about 25% are not eventually licensed. Firms with a higher number of own patents and older firms are more likely to take a license. A licensed patent is looked at on average 3.24 times, compared to 2.23 times for patents that have been considered but never licensed.

Because market *safety* issues are ameliorated in this market, we hypothesize that the lack of demand is due partly to the necessity for complementary technologies in the licensing firm. We measure technology complementarity by utilizing widely recognized technology similarity measures which calculate the overlap of International Patent Classes (IPC) between the AMC patent and the firm's own patent portfolio. We find that technological proximity[1] is indeed a determinant of the decision to in-license once a patent has been looked at. Firms are more likely to license technologies that are similar – i.e. "close" – to their own. While this is true when the proximity measure is computed at the broader subclass level of the IPCs, we note that at the

[1] We use technology proximity and technology similarity interchangeably.

more granular, main group level, conditional on subclass-level proximity, greater similarity between the licensee's patents and the AMC patent makes execution of a license agreement less likely. This implies that "close" fit is good but "very close" fit is detrimental for in-licensing. Additionally, we offer improved measures for technology proximity between patent portfolios.

1.2 Introduction

Markets for technologies (MFT), where ideas and early stage technologies are traded, promise substantial allocative efficiencies and opportunities for productivity growth by promoting gains from trade and specialization of innovative labor (Arora and Gambardella, 2010). They are needed when the locus of innovation is outside of the firm best fit to commercialize it. Suppliers of technology can be lone inventors or users uninterested in entrepreneurship, not-for-profit institutions specializing in publicly funded academic research or firms that do not possess the downstream assets to commercialize their technologies in any or all markets (von Hippel, 1976; Bresnahan and Trajtenberg, 1995; Teece, 1986;). On the demand side, potential efficiencies also exist as firms with downstream assets could use their strengths by buying (better) technology from outside instead of (only) relying on their own R&D capabilities (Pisano, 1990).

The potential benefits of markets for technologies can only be realized if they can efficiently provide *stable* matching between each idea for sale and the firm best fit to commercialize it (Gale and Shapley, 1962; Roth, 2008). Market design theorists have pointed out a few characteristics of markets that are needed for such efficiency – *thickness*, *lack of congestion* and *safety*. A market is *thick* if a large proportion of the potential buyers and sellers participate in the market. It is *not congested* if it gives each participant an opportunity to consider multiple transactions. And, finally, it is *safe* for participants when they choose the market over

other ways of transacting and reveal their true preferences without engaging in welfare reducing strategic behavior (Roth, 2008).

Gans and Stern (2010) highlight the three main characteristics of ideas that can prevent markets for technologies from operating efficiently – *idea complementarity, user reproducibility and value rivalry*. *Idea complementarity* is the notion that ideas are only useful in combination with other complementary ideas. Its existence reduces the number of potential matches to any given buyer or seller and increases the requirements for *market thickness*. *User reproducibility* refers to the fact that once disclosed, ideas can easily be reproduced and the buyer can then become a seller or not pay for the idea (Arrow, 1962). *Value rivalry* is the fact that value gained by one user may diminish as others also use the idea. U*ser reproducibility* and *value rivalry* can reduce *market safety* by inducing strategic behavior by the participants which would result in overall reduction of welfare (Roth, 2008)

Strategy research related to Markets for Technologies has concerned itself mostly with *market safety* issues that may force firms to choose to not transact in the market or can make them engage in strategic behavior (Arrow 1962; Pisano, 1990; Gans et al, 2008; Anton and Yao, 1994; Arora and Fosfurri, 2003; Teece, 1986; Zeckhauser, 1995) In this study we are able to abstract from market safety issues and concentrate on *idea complementarity* and its significance for market *thickness.*

In our paper we explore a small market for technologies in the context of technology licensing from an Academic Medical Center . We observe not only all concluded licenses but also the population of all firms who showed an interest in our sample of patents by signing a confidentiality agreement, evaluative material transfer agreement or an option to an exclusive

license. This allows us to describe the structure of demand in a market for technology, something that has never been accomplished before.

This market is special in that problems of *safety* and *congestion* in markets for technologies are alleviated or non-existent. Our ideas are patented providing a good degree of appropriability and reducing issues of reproducibility by non-licensees. Second, while our seller is interested in generating income its overarching goal in licensing is to see these technologies commercialized and serving the greater good. As a result, it is willing to negotiate with the buyer and price is not the reason why a license is not concluded with a potential buyer. Licensing officer incentives are aligned with the goal of commercialization, not profit maximizing, and significant resources and effort are expended in attempt to commercialize these inventions. Furthermore, the institution is in the business of research and patient care and will not compete with the licensor downstream. As a result, it has no strategic reasons to withhold invention related information from the potential buyer. Additionally, asymmetric information problems, especially with regard to uncertainty regarding the technology quality are attenuated – the inventions come from one of the largest and most respected research institutions in the world.

Given the elimination of many *market safety* and *congestion* issues however, we are still faced with a puzzle: of our sample of 285, approximately half (47%) are never licensed and some 85 (30%) are never even looked at. Of those that are looked at, but not licensed, the first firm to look arrives, on average, 2.75 years after the patent has been filed, or approximately 4-4.5 years after the invention disclosure. Of those that are licensed at least once, first license occurs at 4 years after patent filing or approximately 5.5 years from invention disclosure on average. A patent that has been looked at, but not licensed, gets 2.23 looks, while one that has been licensed has been considered for licensing by 3.24 firms and licensed by 2.02 on average.

In this study we show that even when market safety issues have been substantially alleviated, markets for technologies remain *thin* in the sense that a large number of inventions remain not only unlicensed but also never looked at. This leads us to focus on the importance of *idea complementarity* for the efficient working of these markets. We explore the topic by asking the following research question: "How does technology complementarity affect firm decision to buy a specific idea in markets for technologies?"

We hypothesize that a firm's decision to license a particular invention is dependent on how technologically close its patent portfolio is to the patent under consideration. Using widely accepted measures of technological distance we show that firms license inventions that are close to what they own at the broad level of measurement indicating that idea and asset complementarity are important in their decision making process. However, we also find that controlling for broad level fit, a very close fit at the more granular level of measurement lowers the likelihood of a license due to potentially duplicating in-house efforts.

1.3 Literature Review and Hypothesis Development

1.3.1 Markets for technologies

The volume of trade in markets for technologies has been expanding in recent years. Arora and Gambardella (2010) review recent data from various sources to arrive at a market size of approximately $100 billion globally in 2002 which is about double their earlier estimate of $35-50 billion in the mid-1990s. They also estimate that the market has grown at a higher rate than the average global GDP growth rate in the last two decades (Arora et al. 2001; Arora and Gambardella, 2010, cf. Athreye and Cantwell, 2007; Robbins, 2006). Other survey based studies

point to the increasing importance and rate of out and in-licensing by firms (Sheehan et al, 2004; Zuniga and Guellec, 2008; Lichtenthaler and Ernst, 2007; Tsai and Wang, 2009)

There is some evidence, however, that not all technologies supplied get licensed. Using PatVal survey data Gambardella et al. (2007) show that 11% of firm-owned patents in the sample are licensed but another 7% remain unlicensed even when the firm wants to license them. While there is no information on firm effort in the licensing process, patent quality differences explain the firm's willingness to out-license a particular patent but not whether a license actually occurs. This leads the authors to speculate that it is market and organizational inefficiencies that result in such a licensing shortfall. The result is consistent with other findings that firms are unable to find interested parties with whom to even start negotiations in 75% of the cases in which they want to license and are able to conclude licenses for only 4% of the technologies they wish to license. They often cite high search costs for licensees as the reason (Razgaitis, 2004).

1.3.2 Demand in Markets for Technologies

There is little information regarding firms' demand for outside technologies in the literature. The few available studies are mostly based on survey data on firm practices rather than specific licenses, use different definitions of in-licensing and are difficult to generalize by geography or industry. Using data from a survey on low and medium technology firms from Taiwan, Tsai and Wang (2009) find that 95% of the 753 firms in their sample licensed technology from outside. Rate of in-licensing also appears to differ by country. While attitudes towards in-licensing are similar between Japan and the UK, for example, the incidence of in-licensing is higher in Japan where companies also search more for technology to in-license (Pitkethly, 2001).

Some studies imply passivity on the demand side of these markets and show that the party that initiates the licensing contact is often the supplier (Atuhanegima and Patterson, 1993). Ford (1988), however, provides statistics without a source that claim that 66% of technology sellers and 45% of technology buyers report that the buyer is the one that initiates the technology deal. Those who in-license seem to value the technology that they have acquired. In a survey of firms using university technology, Thursby and Thursby (2004) find that more than half of the respondents use university technology in new product development and 23% note that in-licensed patents from universities were crucial in the development of their products.

A large portion of the research on the demand side of markets for technologies has focused on the firm's decision to "make" or "buy" outside technologies and the factors that influence that decision. Pisano (1990) shows that the firm's choice of external or internal sourcing of R&D depends on considerations of market *safety*, specifically concerns of appropriability and future hold up due to small-numbers bargaining. Firms are more likely to acquire external technology to shorten product development times and gain competitive advantage especially in fragmented IP-regimes (Atuahenegima, 1993; Kurokawa, 1997; Cockburn et. al 2010).

Other studies however show that the success of a strategy of external technology acquisition depends on in-house R&D investment indicating that the two are complements rather than substitutes (Cassiman and Veugelers, 2006; Lowe and Taylor, 1998; Tsai and Wang, 2007). Internal R&D is necessary not only to be able to absorb technologies that the firm has decided to acquire but also to monitor the state of the technology outside the firm's boundaries and evaluate potential technology acquisitions (Rosenberg, 1990; Cohen and Levinthal, 1990; Arora and Gambardella, 1994).

With our study we contribute to this literature by describing the structure of demand in a market for technologies. We use a new dataset of patents from an academic medical center and observe all instances when a firm showed an interest in a technology and its decision to conclude or not a license for that technology later. While the supply side studies have focused on the importance of the product and its attributes to understand this market, our demand-side focused study lets us also explore firm characteristics in the licensing decision. Specifically we are interested in the importance of technology complementarity in firm decision making. We are able to look at complementary technological capabilities in the firm in a very concrete way by observing the patents that the firm already owns and their characteristics. This allows us to answer the question: "Does the technology developed inside the firm influence its decision to acquire a specific outside technology, given interest in the technology."

1.3.3 The Importance of Complementary Technologies

The importance of complementary assets in firms' technology acquisition decisions has been explored before (Teece, 1986; Pisano, 1990). Two studies by Killing (1978) and Caves et al. (1983) look at how in-licensed technologies relate to a firm's current products and capabilities. They provide descriptive statistics on the type of technologies that firms in-license using a convenience sample of 34 licensee companies in the UK and Canada with over 80 licenses in 1974. They find that 22 percent of the licenses were concluded to strengthen the firm's existing products and 70 percent complemented their current capabilities. However, they only rely on licensee survey reports rather than a technology proximity measure and their definition of proximity relates to the products and firms' capabilities rather than the firms' existing technologies.

Little is known about the influence of a firm's technology portfolio in acquiring innovation from outside. Related studies have looked at the importance of technological proximity for firms' diversification decisions. Breschi et al. (2003) find that a firm's diversification decision is path dependent and firms expand into related fields. Building on the resource based view of the firm, Silverman (1999) also shows that firms diversify into areas where their existing technological resources are most relevant. Furthermore, in the context of strategic alliances, firms whose technologies are more similar to their alliance partners' prior to the alliance tend to "absorb capabilities" from their partners (Mowery et al., 1996). In fact technological proximity has been used to quantify spillovers (Jaffe, 1987).

In a recent study, Laursen et al (2010) assume that firms license technologies that are close to what they currently hold and show that firms with a more diverse current portfolio of technology, implying higher "monitoring" and "assimilation" capacity, will license technology that is further away from their current in-house expertise. However, while shedding some light on the importance of absorptive capacity for in-licensing, their study uses a control group of firms that do not license at all in the period under study. This could lead to significant selection problems. Firms that never showed an interest in licensing may be different at some unobserved level. Our sample, in that sense, provides a significantly better way of understanding the relationship between firms' own technology and what they eventually license as we observe both firms that license and those that show an interest but later withdraw from the market.

Based on findings above that firms may be more willing to diversify into technologically closely related areas, we propose the following hypothesis:

H1: Firms are more likely to license inventions that are close to their own technological portfolio*, ceteris paribus.*

We expect that a firm is better able to know about available technology in an area that is closely related to its current knowledge base, reducing search costs for outside inventions. Furthermore, once such inventions are identified, it will be less costly for the firm to correctly evaluate it and assimilate such outside technology into its current portfolio (Cohen and Levinthal, 1990; Arora and Gambardella, 1994). The firm's existing technological capabilities will then help it extract the most value from it (Silverman, 1999). In this study we don't witness a firm's search for new technology since we only observe firms in the "evaluation" stage. Additional data, in terms of commercialization outcomes will let us observe the process of "value extraction" from the firm's current resources as well.

More importantly however, technology similarity is necessary because ideas are often only useful with other ideas (Gans and Stern, 2010). Heller and Eisenberg (1998) argue that especially in biomedical research, inventions are so interdependent that when intellectual property rights are held by different entities, commercialization can effectively be blocked in case of coordination failure. Such idea complementarity makes inventions only relevant to a few buyers which further lowers chances of a match in the marketplace. As such, the existence of complementary ideas evidenced by technology similarity will be crucial in a firm's decision to license an invention.

Licensing ideas complementary to the ones that it already owns can greatly benefit a firm that is developing new products. However, we expect that technologies that are very similar to what the firm owns in the sense that they can be substitutes to in-house developed inventions will not be licensed. Let's assume that the quality of the in-house and the in-licensed technology are similar and perfectly observable to the firm. The firm has already incurred significant costs for its version of the invention and expects to receive the full amount of the future revenue stream. If

it decided to in-license a very similar technology, however, it would most likely pay future royalties to the licensor. As a result, it would choose not to license.

The difficulty of evaluating early stage technologies and the costly transfer of tacit knowledge associated with outside inventions will further lower the chances of a firm licensing even if the quality of the outside invention was better (Polanyi, 1966; von Hippel, 1994; Agarwal, 2006). Furthermore, it is possible that many firms have incentives that reward company scientists for advancing their own technology to the product stage rather than in-licensed technology. Those same scientists are most likely the ones who are evaluating outside technology as well. Behavioral issues such as the so-called "not-invented-here" syndrome which may cause scientists to evaluate outside inventions as inferior to their own have also been pointed out as potential reasons for preferring in-house technologies (Katz and Allen, 1982). This leads us to our second hypothesis:

H2: Firms are less likely to license inventions that are technologically very close (i.e. potential substitutes) to their own technology portfolio*, ceteris paribus.*

We are able to distinguish between H1 and H2 by using an improved version of a widely accepted measure of technological proximity - the cosine, i.e. the uncentered correlation between the technological classes of a focal patent and the firm patent portfolio (Jaffe, 1986). Instead of USPTO patent classes however we use International Patent Classes that have a nested structure and allow us to measure proximity at different levels of granularity. As suggested by previous scholars, we also improve on proximity measures by using all of the IPC codes assigned to a patent rather than the main IPC code (Benner and Waldfogel, 2008).

1.4 Data

1.4.1 Research Setting

The main dataset for our study comes from the technology licensing office (TLO) of a large Academic Medical Center. It contains 285 patents filed and granted from 1980 to 2008 and the associated 307 agreements -- options, confidentiality agreements or licenses -- signed with interested firms for those patents between 1980 and 2010. These patents are the result of employee research and invention. Each employee or affiliate is required to assign to the AMC all rights to all intellectual property developed while at the institution or with funds administered through the institution.

The invention commercialization process starts with an invention disclosure from which a patent is filed which then is licensed through the TLO. When an employee thinks she has developed an invention worth protecting she files an invention disclosure form with the TLO. The invention is then reviewed by a TLO officer with expertise in her subject area who takes on the case. After further consultations with the inventor and further research with respect to the invention's commercialization potential a decision is made on whether to file a patent or release the invention into the public domain. An outside legal firm is then retained to do a patentability search and do the patent filing. As soon as the patent process has been started, the TLO starts looking for potential licensees who will develop their technology further and bring it to market.

There are a few ways in which potential licensees can learn of the invention and these have changed over the years based on new technologies and TLO learning. Brief, non-confidential descriptions of the invention are sent to potential licensees by the case manager after market and industry research. The same description is put on the TLO website where firms can search for it. Firms can also find out about new research results and inventions through research

articles and conference presentations by the inventors, through published patent applications or granted patents and of course through direct contacts with the inventor.

Once a firm decides it is interested in a technology, representatives sign a confidentiality agreement (CDA) which gives them access to the confidential description of the research which often includes the patent application and sometimes the invention disclosure as well.[2] The patent application contains valuable information about the invention and the intellectual property rights (IPR) protection strategy. The signing of a CDA does not involve a fee and does not provide an exclusive right to the technology. In fact CDAs with multiple firms at the same time are common. It does however allow the TLO to know of a firm's interest.

Once a CDA has been signed, the firm may return to explore the technology further and reserve the right to license it by signing an option. Options involve some (albeit minimal) fees and often a requirement that the optionee reimburse unreimbursed past and current patent filing and maintenance costs. Amounts can be negotiated and waived if the firm is cash-restrained which is sometimes true of startups. Options normally last less than a year but can be extended for up to an additional year given the right reasons.

A firm can bypass the option stage and decide to sign a license for an invention. Licenses can be non-exclusive, exclusive in field (e.g. diagnostic uses only, a specific treatment area), exclusive, co-exclusive and end user licenses. Licenses usually, but not always, require the licensing firms to reimburse all or a portion of patent expenses and give them the right to participate in decision making on patent prosecution. Licenses provide revenue to the TLO through a combination of a license issue fee, license maintenance fees, milestones payments,

[2] Note that the American Inventors Protection Act granted the USPTO the right to publish patent applications after 18 months from first filing (priority) date. However, it also gives the right to the applicant to request that the application not be published "but only if the invention has not been and will not be the subject of an application filed in a foreign country that requires publication 18 months after filing (or earlier claimed priority date) or under the Patent Cooperation Treaty" - http://www.uspto.gov/patents/resources/general_info_concerning_patents.jsp - accessed on November 23, 2011

percentage of sublicensing fees and royalties. License terms are quite standard based on the technology type but negotiation and variation are possible. Licenses also require that a firm not shelve the technology it is licensing and in addition to maintenance fees and milestone payments may ask for due diligence reports and other evidence of development efforts.

The TLO works with the licensing firm throughout the life of the patents licensed. An exclusive license can be terminated by the firm for any reason and may be relicensed to another firm for development. Amendments can be signed to change due diligence terms or royalty agreement. Sublicenses may be concluded with firms that will develop the invention further or will sell a product in a different market. The TLO does not routinely terminate licenses but if due diligence milestones are not met has the right to.

It is important to note that the licensor in this process is a non-profit institution that does not have the willingness or the ability to compete downstream with potential licensees. One of the explicitly stated goals of the licensor is to see that the technology serve the greater good by being commercialized. As a result, the licensor is interested in maximizing not only the revenue from a potential licensing deal but also gets utility from seeing the technology brought to market and curing human disease or facilitating further research. TLO officers have incentives aligned with those goals.

As is the case with most TLOs, if a conflict occurs between academic and commercial goals, academic goals take precedence.[3] In fact, in all licenses the hospital retains the right to practice the invention for research and educational purposes. When a government funded

[3] Please see a presentation by an officer from another TLO (MIT) for standard TLO goals, http://web.mit.edu/e-club/www/presentations/tlo.pdf

research is licensed, the government also gets a non-exclusive royalty free (NERF) right to the invention and retains march-in rights in service of the greater good. [4]

What this implies for our dataset is that when a license does not occur, it is almost never due to the parties not being able to reach an acceptable price – i.e. negotiate terms. It is because the potential licensee decided that it was no longer interested in the technology for reasons other than price. This is also seen through qualitative data in the case files - comments by officers about why the potential licensee may not have returned for a license after signing a CDA never list price as the reason. Of course, at the very least, to break even the TLO would like to get reimbursed for incurred patent expenses. The utility of the AMC is subject to the costs of patenting, licensing and infringement and license monitoring.

1.4.2 AMC Data

Our data contains all the solely AMC-owned patents that were filed since 1980, after the Bayh Dole act, and were granted by mid-2008. Patents that are jointly owned with other institutions such as universities or companies were excluded as licensing or development activity by the co-owner is not observed in such cases. Patents that are the result of for-profit sponsored research are similarly excluded because sponsorship by industry almost always results in an automatic exclusive option to all the patent rights and in some cases an automatic license with pre-agreed terms which takes the patent off the market. If the sponsoring company did not desire a license after the invention disclosure was made, then that is a signal to other potential licensees that the new technology may not be of high quality since the sponsor would have better private information than a potential licensee that was not involved. In either situation, these patents

[4] http://www.uspto.gov/web/offices/pac/mpep/documents/appxl_35_U_S_C_203.htm - accessed November 24, 2011

would not have been directly comparable to the rest and useful for our analyses. This leaves us with 285 patents available for licensing.

The data also includes all the agreements that were signed with for-profit institutions that have the ability to commercialize an invention protected by a patent. These include confidentiality agreements (CDAs) which indicate that a company has shown interest in a certain patent, material transfer agreements (MTAs) for evaluation of biomaterials or prototypes, options to a license for a certain period of time, end-user licenses, non-exclusive licenses, exclusive licenses, sublicenses and patent assignments.

A unique agreement per company and patent was selected if multiple agreements were signed within 5 years of the first agreement with the same company. For example, if a company signed a confidentiality agreement, then followed up with an option and finally signed an exclusive license for the same patent, only the last was selected. If a company signed a CDA but not a license and then 6 years later signed a license, we included both the original CDA and the license as separate agreements. Similarly, amendments to agreements were not included unless they included additional patents and in that case were only included for the new patent.

For our purposes, agreements were divided into two categories – "deals done" and "deals not done." If a license was signed, the agreement was considered a "deal done" and this included exclusive and nonexclusive licenses and sublicenses. Agreements were classified as "deals not done" if they indicated an interest in the patent through a CDA, MTA or option but no license was concluded. This resulted in 307 agreements and overall 600 patent–agreement pairs since many agreements have multiple patents under them and many patents have been looked at and licensed multiple times.

1.4.3 Firm Data

Each firm's technology profile at the time of agreement signing was compiled using patent data. We matched licensee names to patent assignee names conducting assignee name disambiguation by manually going through more than 450 000 company names and by searching for common misspellings. This is important as company names are not standardized at the USPTO. For example, Microsoft patents can be under Microsoft, Microsoft Inc., Microsoft Inc, Micosoft (misspelling) and direct matching to Microsoft Inc. would exclude multiple patents under the other names. In addition, certain companies patent under subsidiary names. For example Zeneca Plant Science and Zeneca Pharmaceuticals are part of the same company and we assume that patents assigned to Zeneca Plant Science are also available for use to Zeneca Pharmaceuticals and vice versa without a license.

Alternative automated ways for name matching are available and we used the 'soundex' function that assigns a string consisting of a letter and numbers to words based on how they sound. We also tried matching using the SAS 'COMPGED' function which measures the "edit distance" between two strings – i.e. the number of deletions, insertions, or replacements in the characters of a word required to arrive at the observed word.[5] In our case, both were inferior to manual matching as they excluded many relevant observations and included irrelevant ones.

For the purposes of this paper, we defined a company's technology position as a stock of patents filed before the time of agreement signing and did not use a depreciation factor for older patents. There are a few studies that show that learning depreciates over time (e.g. Benkard, 2000) and company focus may change and lead to a different technological expertise now from

[5] http://support.sas.com/documentation/cdl/en/lefunctionsref/63354/HTML/default/viewer.htm#p1r4l9jwgatggtn1ko81fyjys4s7.htm accessed October 1, 2011.

the one many years ago. However, it is not clear how long it takes for such technological expertise to change or expire.

One way to determine how long certain technological expertise is relevant for the company is to look at patent validity – if the patent is still valid, then the firm still has that technological ability. However, data limitations prevent us from finding out what patents are still 'alive'. Computation of patent validity at a point in time is impossible without the availability of patent priority data which determines patent term. We have manually gathered that data for AMC patents but we don't have such data for firm patents. In addition, not all patents are maintained to the end of their term. The assignee needs to pay a certain fee to keep a patent alive 3.5, 7.5 and 11.5 years after it is granted.[6] Such payment data is also not available in an aggregated form. Furthermore, some patent terms are adjusted because it takes the USPTO longer to review them. That information is not available in an aggregated form, either.

One difficulty with determining patent portfolio size and content for the companies that are party to these agreements is that most of them are in the pharmaceutical and biotech industry which have seen many mergers and acquisitions in the last few decades. We could not find data on such M&A activity until 1992 and the post-1992 data is not complete for all of the companies so was not included in this version of the paper.[7] The available data is often difficult to interpret as many companies sell or acquire specific plants or businesses such as the "vaccine business" or their "nutritional business" but it is unclear what patents are licensed or sold off with these divestitures and acquisitions. The only exceptions to the exclusion of M&As from our dataset are the top 10 pharmaceutical companies that have seen multiple large mergers and acquisitions in

[6] This term can be extended to 4, 8 and 12 years with the payment of a fine in the six months between 3.5th and 4th anniversary.

[7] For post-1992 we were able to find some data through Lexis-Nexis Company Dossier service but the data was not complete, especially for non-US firms

the 1990s – 2000s. Acquisitions of large companies or mergers between two or more large companies were tracked using the web, mainly through company history pages, and verified through Lexis-Nexis Company Dossier service for M&As after 1992. Examples of these include the acquisition of Hoechst by Aventis, the merger of Glaxo Wellcome with SmithKline Beecham and so on.

It is important to note at this point that four companies and two joint ventures with eight agreements and twelve patent-agreement observations were excluded because the company data was not reliable. Regarding the two JVs, it was not clear what kind of parent company patents and knowledge they had access to. Considering them completely unrelated to the parent companies was probably not correct either. All 4 of the remaining companies are highly diversified, with over 30,000 patents each and two of them have a pharmaceutical or medical device business but the majority of their business is in other industries such as electronics, manufacturing, household goods, aviation, and finance. One of these companies was so large that it had completed 982 acquisitions and divestitures in the 18 years since 1992. Another one had a very common name and it was not possible to disambiguate the name from the other companies in the patent assignee file with the limited resources we had. By mistakenly including patents of unrelated companies, we would mischaracterize the licensee's technological position. Including these diversified companies would have made them not comparable to the other companies and would have distorted our results. This brought down our sample size to 295 agreements and 588 patent-agreement observations.

1.4.4 Patent data

We use patent data to characterize the technology position of a company and the specific AMC technology (Jaffe, 1987; Silverman, 1999). It has been noted that patents are an

incomplete indicator of a firm's stock of knowledge as they don't account for expertise that is not protected by a patent. However, multiple studies have used patent data as a proxy for a company's knowledge base using the argument that measures based on knowledge protected by a patent are highly correlated with uncodified and unpatented knowledge (Silverman, 1999; Patel and Pavitt, 1994; Narin et. al., 1987).

Patent data have an advantage in that they are reliably available since 1963 and contain information that can be used to characterize the specific inventions in various ways. The most important information for our analyses is the technology class which has been assigned to the patent. Unlike many other papers that have relied on the USPTO classification system, we use the International Patent Classification (IPC) codes assigned to our patents by the EPO as found in the PatStat database.[8] We also use all of the patent IPC codes rather than assigning the first one to be the 'main' IPC code. Since the EPO itself does not assign a main IPC code we do not worry about weighting one IPC code more than another (Benner and Weldfogel, 2008).

The IPC codes are different from the USPTO classification codes in multiple ways. Most importantly, each IPC code has five nested levels of detail from broad to detailed – section, class, subclass, main group and subgroup level. USPTO classification provides only a class and a subclass. Furthermore, IPC codes divide the technology spectrum into finer slices - at the IPC subclass level which is comparable in the level of detail to the USPTO class level: there are 640 unique subclasses while the USPTO has only about 400 unique classes (Hall et al. 2001, WIPO website[9]). This nested quality is important for our analyses because it lets us measure technology fit at various levels of detail. Furthermore, with the exception of the finest level of

[8] Pat Stat website at: http://www.epo.org/searching/subscription/raw/product-14-24.html, accessed January 2012

[9] WIPO website, FAQ, http://www.wipo.int/classifications/ipc/en/faq/index.html accessed Sept 9, 2011

measurement – the subgroup level (which we don't use) - patent IPC codes are not laterally nested.[10] USPTO patent classes are often laterally nested at the subclass level.

Another important piece of information that is available on a patent are the citations a patent makes to other patents and the citations future patents make to a focal patent. In addition to providing links to other patents such citations are used to determine whether a patent represents a pioneering or more incremental invention. Unlike citations in academic articles, the citations to previous patents, also referred to as *prior art*, delineate and limit the scope of a patent. If a patent cites a prior patent, it means that it cannot lay a claim to the invention in the previous patent. Normally, the more prior art a patent has, the more incremental it is considered and the more developed the technological area to which it belongs. Conversely, fewer backward citations imply that a patent is pioneering. Similarly, forward citations are used to determine the importance of the patent and are correlated with the value of inventions (Trajtenberg, 1990). If a patent is cited by numerous patents, it indicates that many inventors are building on the original invention and hence the invention is more significant (Hall et al., 2001). Furthermore, while only authors can add citations to papers, patent citations are added to the patent both by the patent filer and the patent examiner.

In our paper we use count of patents cited by the focal patent as one of our explanatory variables. The use of forward citations of the focal patent, however, presents some difficulties. To receive forward citations a patent has to be published and the longer a patent has been published the higher chance it has of being cited (Mehta et al., 2001). Through analysis of the

[10] What this implies is that two patent IPC codes that appear next to each other are not nested within each other at t any of the levels except the subgroup level, which we do not use. For example A61F 2/04 is a subset of A61F 2/02 at the subgroup level. However, A61F 2 and A61F 3 (at the main group level) are not nested within each other. See at http://www.wipo.int/ipcpub/#refresh=symbol&notion=scheme&version=20120101&symbol=A61F0002940000, accessed January 11, 2012.

overall US patent data Hall et al. (2001) show, that the average patent receives just one half of its lifetime citations by the tenth anniversary of its publication. About 48% of our hospital patents have been granted since 2000 and as such have likely not received even 50% of their expected lifetime citations (Hall et al., 2001). Our data is right censored - the last patent in our dataset was granted in July 2008 and has had only two years to get cited. Because we can't compare total citation count, we construct a variable that measures 'citations per year.' The cites per year variable makes patents of different ages comparable with regards to their forward citations but at the expense of making a strong assumption that the distribution of forward citations is uniform with regard to time. It would, for example, underestimate the importance of younger patents that receive fewer citations early but become very important later on.

A combination of patent citation and patent technology class information has led to the creation of other, composite measures to describe patents. Two important ones are patent *originality* and patent *generality*. The patent originality measure uses the sum of the squared shares of cited patents that belong to a certain technology class to create a Herfindahl-Hirschman index (HHI) over all technology classes. The measure is 1 minus that HHI. Since we use multiple technology classes per patent, our measure is modified to include not the share of *patents* in each class but the share of *patent IPC codes* assigned to the cited or citing patents. The fewer different IPC codes the cited patents belong to, the smaller the originality measure and less original the patent is considered to be. For example, if a patent that belongs to organic chemistry cites patents that are in software, organic chemistry and aviation, that would be considered a more original patent than one that cites only patents in organic chemistry (Hall et al., 2001).[11]

The generality measure is created in a similar way – only this time, forward cites are used and the HHI is created over the squared share of each IPC code summed over all citing patent

[11] We used the correction in Hall et al., 2001 for the calculation of these measures.

IPC codes and subtracted from one. A focal patent that is prior art to patents in multiple different classes is considered more general as it is a platform for multiple technologies in different fields. An example of a very general patent here could be an invention in molecular biology that has been cited by patents in classes as diverse as plants, drugs, electricity, manufacturing and electronics. Compare that to a less general molecular biology patent that is only cited by other patents in the molecular biology field (Hall et al., 2001). The generality measure has a serious shortcoming because most patents have not received all of their future cites at the time of observations and hence their generality may be underrepresented.

Another composite measure created from patent statistics is the patent scope. It is defined as the number of IPC codes assigned to a patent. The more IPC codes a patent has, the broader its scope is considered to be. It has been shown to determine patent valuations by venture capitalists and patent licensing outcomes (Lerner, 1994; Gambardella et. al., 2007; Decheneaux et al., 2008)

1.4.4.1 Technology Similarity/Proximity Measures

The main independent variables that we construct using patent data are our technology proximity measures between the focal AMC patent(s) and the patent portfolio of the interested firm. For this purpose we use the cosine measure pioneered by Jaffe (1986) but with modifications that utilize all IPC codes instead of the main USPTO patent class and at different levels of detail of the IPC code - subclass and main group.

The cosine measure calculates the angular distance between two vectors that characterize the firm's and the AMC patent's position in a technology space defined by patent classes. For this purpose we create a technology position vector for a firm's portfolio $F_i=(F_{i1}, F_{i2}, F_{i3}\ldots F_{ik})$, where each 'entry' is the share of a firm's patents (in our case IPC codes since one patent may

have multiple IPC codes) in a certain technology class k. A technology position vector, F_j is also created for the specific hospital patent under the firm's agreement. The angular distance between the two vectors is then the measure of technology similarity and it ranges between zero and one, one being a perfect fit and zero being no overlap in technology. It is calculated using the following formula:

$$P_{ij} = \frac{F_i' F_j}{\sqrt{(F_i' F_i)(F_j' F_j)}}, 0 \leq P_{ij} \leq 1$$

For companies that did not have patents filed before the date of the agreement the cosine measure is not defined. We replace the proximity measure for such observations with zeros indicating no fit and run results with and without replacement. We also provide descriptive statistics and models where we exclude observations for which the cosine measure is not defined.

We modify the cosine measure further by computing a "within section" cosine measure constructed based on the above formula, except we exclude all IPC codes in the firm patent portfolio which do not match the AMC focal patent IPC codes at the section levels. For example, if the AMC patent under consideration has the following two IPC codes A61K 9/12 and C07B 12/07, we only keep the firm IPC codes that are in the A and C sections deleting IPC codes in B, F, G and H. This measure looks at proximity of the closest part of a company's technology portfolio to the focal patent. Again firms that are left with no IPC codes that are in the two sections will get a cosine measure of zero.

Both the broad and the within section cosine measures are calculated at the group and subclass IPC code level. However, in our models we use the regular cosine measure at the broader level and the within section cosine measure at the more detailed, main group, level. Our results do not change in direction or magnitude if we replace the within-section cosine measure

at the main group level with the regular cosine measure at the main group level (i.e. without excluding the sections that do not match the AMC patent IPC code sections)

1.4.5 Control Variables

For each patent – agreement pair, we also calculate a technology age measure which is the time in years from the hospital patent's patent priority date to the agreement date. The priority date is the date on which the first patent on a certain invention disclosure was filed with the USPTO. This original patent can then be divided into multiple patents if the USPTO deems that it contains more than one separate invention. Continuation patents can be filed from the original patent and continuation-in-part patents, in particular, can add some new matter. [12] Since the priority date is closest to the invention disclosure date, technology age calculated with the priority date rather than patent filing date is an indication of how mature the technology is. Technology risk is more likely to be resolved for older inventions as inventors or others have developed it further to bring it closer to market.

An indicator variable for device was generated by looking through each patent's claims. Claims of an apparatus with human body contact or some sort of an implant were marked as devices. The criterion was whether the invention would have required an approval by the FDA as a device in order to be used in the market. For example, an apparatus for growing cells would not be considered a device but rather a research tool. Devices include stents, artificial joints, catheters, surgical instruments, MRI machines and so on.

To describe the technology further, we used indicator variables for the IPC code sections to which the patent belongs. Since many patents have multiple IPC codes that sometimes belong

[12] Note that the priority date is NOT the same date as a provisional patent date. For more information on priority dates, please see http://www.yale.edu/ocr/pfg/guidelines/patent/continuation.html, accessed November 29, 2011

to different sections, we made them into mutually exclusive categories i.e. patents in section A only, patents in section A and C only and so on. Most of our patents are in section A – "Human Necessities" which includes class A61 – "Health; Life Saving; Amusement" which includes most drugs and medical devices. The next important section in our data is section C – "Chemistry; Metallurgy" which under classes C07 and C12 includes organic chemistry and biochemistry, molecular biology and genetic engineering. Section G – "Physics" includes patents related to imaging, ultrasound and measuring. The omitted category in our models is "C&G only" with 22 observations and 4 additional patents that include IPC codes in B or F section.

We also compute a few firm level variables based on patent measures. Because many of our firms, especially those that appear multiple times in our data undergo mergers and acquisitions, we are not able to include firm fixed effects. Furthermore, even though some firms have multiple agreements, most firms have only one or two. Our main concern in controlling for firm differences was to separate the old pharmaceutical companies whose expertise is mostly in small molecule drugs from the biotechnology companies that are specialized in molecular biology. To be able to do this, we created two variables – R&D age at time of agreement and total number of granted patents applied for before time of agreement. The R&D age is defined as the time from the first filing of a patent by the firm to the time that it signs the specific agreement with the AMC. Unfortunately, our patent data goes only to 1963 so big pharmaceutical firms that have been in existence since the 1800s and licensed a technology in 1990 will be at a similar age with a firm that was created in 1980 and licensed a technology in 2007. Similarly R&D age is not defined for firms with no patents.

Another firm level variable that is better at distinguishing the old pharmaceutical companies from the young biotechnology companies is the number of firm patents at the time of

agreement. Since old pharmaceutical companies have undergone multiple mergers, they have a substantially larger number of patents than young biotech firms. In that sense the "number of patents" variable is a proxy for both size and age. Because there is a high correlation between R&D age and number of patents, we only include the number of patents variable in our models. Because we expect non-linear effects as well, since this variable's range is very high, we also use its square in our analyses.

We are interested in how the influence of proximity on licensing varies with age and size – i.e. with the type of the firm. Because we have two proximity measures, interaction effects will be difficult to interpret. For that reason, we separate our firms in three groups based on age – early startups, growth startups and old companies where early startups will have filed their first patent between 0 and 10 years before an agreement was signed with our AMC. Mature startups will have filed their first patent between 10 and 20 years before the agreement and old firms would have filed their first patent more than 20 years before the agreement. Based on size, we split the firms in two groups – those that have fewer than 500 patents and those that have more than 500 patents. Those that have fewer than 500 patents are generally biotech or medical device firms. The other group contains all the small molecule pharmaceutical companies.

1.4.6 Descriptive Statistics

1.4.6.1 Patent Level Descriptive Statistics

Descriptive statistics for the patent level dataset are included in Tables 1.1a and 1.1b. The patents are separated into three different groups – those that were never looked at, those that were looked at but were never licensed and those patents that were licensed at least once. Significance levels of two tailed t-tests of comparisons between the first two groups and the "licensed" group are indicated next to the mean of the variable in the respective group. For

Table 1.1a This table contains descriptive statistics for the AMC patents that are to be licensed. Two sided t-tests of difference in means between "Never Looked At" and "Licensed at Least Once"; significance indicated in the "Never Looked At" mean column; Two sided t-test of difference in means between "Looked At, Not Licensed" and "Licensed At Least Once"; significance indicated in the "Looked At, Not Licensed" mean column; *** p<0.01, ** p<0.05, * p<0.1

	Never Looked At					Looked At, Not Licensed					Licensed At Least Once				
	N	Mean	Std Dev	Min	Max	N	Mean	Std Dev	Min	Max	N	Mean	Std Dev	Min	Max
Lead Inventor Experience	**85**	4.61***	3.43	1	15	**48**	6.02	4.61	1	17	**152**	7.19	4.77	1	25
Patent Scope	**85**	2.43***	1.38	1	9	**48**	2.87	1.48	1	6	**152**	3.71	3.49	1	19
Cites Per Year	**85**	0.78	1.18	0	6	**48**	0.33**	0.59	0	3	**152**	0.96	1.83	0	19
Cites First Two Years	**85**	0.42	0.90	0	6	**48**	0.33	0.88	0	4	**152**	0.47	0.85	0	5
Originality	**85**	0.64	0.32	0	1	**48**	0.67	0.29	0	1	**152**	0.68	0.23	0	1
Generality	**62**	0.62	0.20	0	1	**25**	0.72***	0.22	0	1	**126**	0.57	0.24	0	1
Number of Cited Patents	**85**	6.57***	6.97	0	29	**48**	6.83**	7.27	0	37	**152**	10.16	10.19	0	54
Share of Agreements that are Licenses											**152**	0.79	0.28	0.125	1
Times Looked						**48**	2.22**	1.85	1	8	**152**	3.24	3.12	1	15
Times Licensed											**152**	2.02	1.66	1	11
Time to First Agreement						**48**	2.75	2.45	0.00	7.98	**152**	3.44	3.68	0.00	12.76
Time to First License											**152**	4.20	3.84	0.00	18.48
Age of Oldest Cited Patent	**66**	15.69**	11.48	2.34	55.56	**45**	17.29	12.66	1.36	57.55	**140**	21.55	17.34	1.09	74.81
Mean Age of Cited Patents	**66**	7.53***	3.95	1.94	18.73	**45**	8.71	4.31	1.36	16.99	**140**	9.94	4.84	1.09	25.21
Median Age of Cited Patents	**66**	6.38***	3.52	1.74	14.46	**45**	7.74	3.92	1.36	16.99	**140**	8.68	4.25	1.09	22.84

Table 1.1b This table contains descriptive statistics for the AMC patents that are to be licensed. Two sided t-tests of difference in means between "Never Looked At" and "Licensed at Least Once"; significance indicated in the "Never Looked At" column; Two sided t-test of difference in means between "Looked At, Not Licensed" and "Licensed At Least Once"; significance indicated in the "Looked At, Not Licensed" column; *** p<0.01, ** p<0.05, * p<0.1

Group	Never Looked at (N=85)	Looked At, Not Licensed (N=48)	Licensed (N=152)
	Proportion of "Never Looked At" Patents in Group	Proportion "Looked At, Not Licensed" Patents in Group	Proportion of "Licensed" Patents in Group
IPCs in Section A Only	0.36	0.35	0.38
IPCs in Section C Only	0.14	0.06	0.14
IPCs in Section G Only	0.22***	0.14**	0.05
IPCs in Sections A and G Only	0.02	0.04	0.05
IPCs in Sections A and C Only	0.18*	0.19	0.30
IPCs in Sections A, C and G Only	0.01*	0.06	0.07
Patent Priority Date in 1981-85	0.07	0.00*	0.07
Patent Priority Date in 1986-90	0.28	0.02***	0.25
Patent Priority Date in 1991-95	0.23***	0.08***	0.43
Patent Priority Date in 1996-00	0.30*	0.66***	0.20
Patent Priority Date in 2001-04	0.10*	0.22***	0.05
Device	0.38***	0.29	0.20

example, the stars next to the mean value of the variable “Number of Cited Patents” in the “Never Looked At” group indicate that the difference of the means of the “Number of Cited Patents” between the “Never Looked At” and “Licensed At Least Once” groups is statistically significant.

The cohort variables in Table 1b that group patents into cohorts based on their priority date show an interesting story– 88% of the patents that were looked at but not licensed have a priority date after 1995 compared to only 40% of those that have never been looked at and 25% of those that have been licensed. It is important to point out this can influence how the rest of the variables are distributed across group. This observation is not surprising because we expect that patents that have been only recently filed pose a higher risk for commercialization as the technology has not yet been tested and proven. We expect that the probability of licensing increases up to a certain patent age and then decreases as the remaining patent life becomes too short for high investments to be recovered.

The “lead inventor experience” variable is defined as the number of inventions that the inventor has previously disclosed and patented at this specific technology licensing office (TLO). It is a proxy for the inventor’s experience both innovating and navigating the licensing process. Many firms can view an inventors’ previous experience as a signal of the quality of the invention and its commercialization potential. Furthermore, a larger number of inventions can imply stronger and broader IP rights if the inventor has worked on similar problems before and his previous inventions are related to the current ones. We note that, as expected, the “lead inventor experience” variable is smaller in the “never looked at” group than in the licensed group and this difference is statistically significant. The patent scope variable, defined as the number of unique

IPC classes on each patent (Lerner, 1994) is also intended to measure breadth of IP rights. In this case we find that it differs significantly between the “never looked at” and “licensed” group.

As expected, forward cites per year and two year forward cites are largest among the licensed patents indicating that those are more important and more inventions build on them. However, the result is only statistically significant for the difference between the “Looked At, Not Licensed” group and the “Licensed” group.

“Number of Cited Patents” is an indicator of how pioneering the technology is. Radically innovative patents would have little or no prior art because if a patent cites a previous patent, it means that it builds on it. We note that while 92% (140 out of 152) patents in the “Licensed” group cite prior patents, only 77% (66 out of 85) do in the “Never Looked At” group. This indicates that the more pioneering a patent is, the less likely it is to be looked at or licensed and the trend holds over all groups. This also implies that university technology is ahead of industry developments.

This implication that firms prefer to license in more established technologies even if that leads to narrower IP rights is confirmed by the average age of patents cited by the AMC patents under the agreements. We note that of those patents that cite at least one prior work those in the “Never Looked At” group cite on average younger patents than those in the other groups and this difference is statistically significant.

Patent originality and generality use the IPC classes to which forward cites or backward cites belong. As such when patents have no prior art or no forward cites as of yet, their generality and originality measures are undefined. As a result there are fewer observations in these categories. For originality, we substitute one which equals fully original for patents that have no

prior art. While it does not follow the formula by which originality is calculated, it follows the intuitive understanding of originality.

Our generality measure depends highly on the age of the patent. Patents that are younger may not even have a generality measure that is defined because they don't have citations. Older patents, on the other hand, are expected to be on average more general as they have more citations which will have a higher chance of belonging to multiple unique classes. In fact, in our dataset we see that while only 25% of the whole set of patents are missing a forward citation and a generality measure, only 50% of the 'looked but not licensed category', with the youngest patents on average, have any forward cites. And while we see that as a whole they have the highest mean generality measure, this may reflect the influence of outliers in a smaller sample. We can't replace the generality measure with a zero for patents that have no future citations because it is possible that such citations will be received in the future. As such, we don't use the generality measure in most of our models since it significantly reduces our sample size.

Interestingly, while 38% of the patents in the 'never looked at' group are devices, only 29% and 20% in the 'looked at but not licensed' and 'licensed' categories respectively are devices, indicating that devices are much less likely to be licensed. Licensed patents also have a higher number of agreements associated with them – on average 3.24 compared to only 2.23 for the 'looked at but never licensed' group and this finding is statistically significant.

1.4.6.2 Patent-Agreement Level Descriptive Statistics

The next level of descriptive statistics is at the agreement – patent level. In this dataset there are 588 observations and each observation is a patent agreement pair. Each agreement is either a 'deal done' or 'deal not done' and some agreements contain multiple patents. Since some

Table 1.2a This table contains descriptive statistics for the agreement-patent level data. Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. Patent level measures correspond to the hospital patent which is under the agreement. Two sided t-tests of difference in means between "Deal " and "No Deal" ; significance indicated in the "No Deal" mean column; *** p<0.01, ** p<0.05, * p<0.1

Variable	No Deal					Deal				
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Cosine Subclass Level†	287	0.24**	0.31	0	1	301	0.30	0.31	0	1
Cosine Subclass Level (no replacement)††	198	0.35	0.32	0	1	226	0.39	0.30	0	1
Within Section Cosine, Group Level†	287	0.22	0.31	0	1	301	0.22	0.28	0	1
Within Section Cosine, Group Level (no replacement)††	198	0.33	0.32	0	1	226	0.29	0.28	0	1
Patent Scope	287	3.06*	1.75	1	10	301	3.45	2.94	1	19
Patent Citations Received Per Year	287	0.57***	0.74	0	3.67	301	0.95	1.79	0	18.91
Patent Citations First Two Years	287	0.38	0.74	0	5	301	0.45	0.87	0	5
Number of Patents Cited	287	10.88	10.66	0	54	301	10.53	10.22	0	54
Patent Originality	287	0.71	0.22	0	1	301	0.71	0.20	0	1
Patent Generality	194	0.60	0.26	0	1	246	0.58	0.24	0	1
Device	287	0.15	0.36	0	1	301	0.12	0.33	0	1
Technology Age in Years	287	5.02***	4.38	0	18.11	301	6.64	4.75	0	19.35
Lead Inventor Experience	287	7.85	5.10	1	22	301	7.27	4.52	1	25
Firm R&D age	198	17.04***	11.47	0.06	43.27	226	21.20	12.01	0.08	46.17
Number of Firm Patents	287	985.01***	2657.38	0	16392	301	2204.26	4227.66	0	15266
Numbers of AMC Patents Under Agreement	287	2.75***	1.70	1	6	301	4.20	4.36	1	19

†Observations for which a cosine measure was not defined because the firm has no patents of its own were assigned a cosine measure of 0.

††Observations for which a cosine measure was not defined were excluded from this calculation.

Table 1.2b This table contains descriptive statistics for the agreement-patent level data. Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. Patent level measures correspond to the hospital patent which is under the agreement. Two sided t-tests of difference in means between "Deal " and "No Deal" ; significance indicated in the "No Deal" mean column; *** p<0.01, ** p<0.05, * p<0.1

	No Deal (N=287)	Deal (N=301)
Group	Proportion of "No Deal" Patent Observations in Group	Proportion of "Deal" Patent Observations in Group
Firm Has No Patents/Firm R&D age undefined	0.31*	0.25
Firm Has >0 and < 500 Patents	0.47	0.43
Firm Has >500 Patents	0.22***	0.32
Firm R&D age >0 and < 10 yrs Old	0.25***	0.13
Firm R&D age >10 and <20 yrs	0.18**	0.27
Firm R&D age >20 yrs	0.26**	0.35
Patent Priority Date in 1981-85	0.13	0.13
Patent Priority Date in 1986-90	0.13***	0.25
Patent Priority Date in 1991-95	0.09***	0.36
Patent Priority Date in 1996-00	0.53***	0.25
Patent Priority Date in 2001-04	0.13***	0.02
IPCs in Section A Only	0.42	0.40
IPCs in Section C Only	0.07***	0.15
IPCs in Section G Only	0.05	0.03
IPCs in Sections A and C Only	0.25	0.31
IPCs in Sections A and G Only	0.04	0.06
IPCs in Sections A, C and G Only	0.10**	0.05

patents also have multiple agreements on them, a patent may be in the 'deal not done' column with a certain agreement and in the 'deal done' column with a different agreement.

Our most important independent variables are the proximity measures. The cosine measure at the subclass level defines the overall broad-level proximity between a firm and a patent while at the main group level it represents proximity at a finer detail. The main group level cosine measure is more appropriate for "within section" fit. Both measures have a range between 0 and 1 with 1 indicating a perfect similarity and 0 indicating no similarity between the focal patent and the firm patent portfolio. We note that our cosine measure at the subclass level is smaller in the "no deal" group than in the "deal" group, indicating that firms license technologies that are closer to the technology that they own. At the cosine main group level measured within the AMC patents' sections, there seems to be no difference between the two groups.

It is important to note that the cosine measures are not defined for agreements with firms that have no patents as they don't have IPC classes for matching with the AMC patent. In Table 1.2a, we first show the mean of this variable after we replace the cosine measure with a zero indicating no similarity between the patent and firm technology for firms that have no patents. The difference of the means of the cosine subclass level between the "deal" and "no deal" groups is statistically significant in this sample. We then exclude those observations where the cosine is not defined and calculate the means without replacement. The means of the cosines are not significantly different between the groups any more.

Firm R&D age differs significantly between the two groups, with older firms more likely to conclude licenses. In fact, it is firms that are younger than 10 years that drive this result as seen in Table 1.2b. They are likely to look at patents but not license. This is also true for firms that have no patents which could be very young i.e. startups but which could also be firms that

are not engaged in R&D such as contract research organizations (CROs). Similarly, firms with more patents are more likely to license. The results are also confirmed by the age and size distribution of firms in the "deal" and "no deal" groups in Table 1.2b. In fact, Table 1.2b shows that a larger share of the "no deal" than of "deal" observations are with firms that have no patents or have fewer than 500 patents. Finally, licensing agreements ("deals") include more hospital patents than "no deal" agreements, indicating that firms prefer to license large portfolios of patents indicating more complete intellectual property rights for related inventions.

In confirmation of our previous result on the age of technology offered for licensing, we find that "deals" are significantly more likely to occur when the technology is older, indicating again that firms prefer more established inventions. They also prefer inventions with higher impact indicated by the difference between means in the two groups along the "citations received" variable. As expected from the patent dataset devices represent a higher percentage of the "deals not done."

The last set of descriptive statistics is at the agreement level. The difference from the previous table is that we are counting each agreement only once. We started with 307 agreements and excluded 12 firms for a sample size of 295 agreements. Again we report cosine statistics with replacement and no replacement. The cosine statistics at the agreement level are calculated over all the IPC codes over all the patents that the hospital has under the agreement. Also, at the agreement level we do not report a within section cosine, but rather a regular cosine at the group level. The results are in the same direction as the ones at the patent-agreement level but the differences in means between the two groups are not significant any more.

Table 1.3 This table contains descriptive statistics for the agreement level data. Each observation corresponds to a unique agreement. Two sided t-tests of difference in means between "Deal " and "No Deal" ; significance indicated in the "No Deal" mean column; *** p<0.01, ** p<0.05, * p<0.1

	No Deal					Deal				
Variable	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Number of Firm Patents	153	1028.43	2740.24	0	16392	142	1328.67	3039.97	0	15266
Firm Technoology Age At Time of Agreement	153	12.44	12.40	0	43	142	13.33	12.79	0	46
Number of Patents on Agreements	153	1.88	1.28	1	6	142	2.12	2.11	1	19
Agreement Level Cosine, Main Group Level†	153	0.20	0.27	0	0.96	142	0.18	0.26	0	0.94
Agreement Level Cosine, Main Group Level (no replacement)††	114	0.26	0.28	0	0.96	99	0.26	0.27	0	0.94
Agreement Level Cosine, Subclass Level†	153	0.27	0.32	0	1	142	0.29	0.33	0	1
Agreement Level Cosine, Subclass Level (no replacement)††	114	0.36	0.33	0	1	99	0.42	0.32	0	1

†Observations for which a cosine measure was not defined because the firm has no patents of its own were assigned a cosine measure of 0.

††Observations for which a cosine measure was not defined were excluded from this calculation.

1.5 Results:

1.5.1 Patent Level Models

Because we have some patents that are “never looked at”, some that are “looked at but never licensed” and some that are “licensed at least once”, we run a number of analyses to understand what patent characteristics may influence licensing or interest in an invention in general. We run a multitude of regressions using the variables that we have described above. The results are in Tables 1.4a and 1.4b.

The first models we run, reported in Table 1.4a try to explain what variables influence whether a patent has been looked at by a firm, how many times it has been looked at and how much time has passed before the first look. Each agreement, whether a CDA, MTA, option to license or a license is a “look”. The second set of models, in Table 1.4b has licensing outcomes as the dependent variable – whether a patent has been licensed, how many times it has been licensed and time to first license. The first models in Table 1.4a and 1.4b are logit models where the dependent variable is whether a patent has an agreement on it i.e. has been looked (Table 1.4a) and whether a patent has a license on (Table 1.4b). Interestingly, we see that the variable that measures the number of citations received, an indication of importance, is not statistically significant in explaining whether a patent is looked at or licensed. However, the number of previous cites is positively related to whether a patent is looked at or licensed, indicating that more established technologies are more likely to be successful in markets for technologies. We also note the significance of the lead inventor experience pointing to the importance of quality signals in a market with a lot of product uncertainty

Similar to the descriptive statistics, we see that patents that are devices are less likely to be successful. This is unexpected given the importance of physicians in new medical device

Table 1.4a This table contains models of patent consideration for licensing i.e. "looks at patent" based on patent characteristics. P-values reported under coefficients.

	LOGIT	**Poisson**	**Negative Binomial Regression**	**Zero Inflated Poisson**		**Cox Hazard Model**	**Cox Hazard Model** Stratified by Lead Inventor	**Cox Hazard Model** Frailty by Lead Inventor
VARIABLES	Looked (looked=1)	Times looked	Times looked	Times looked	Inflate (looked=0)	Time to Look	Time to Look	Time to Look
Number of Patents Cited	0.0837***	0.0311***	0.0346***	0.0225***	-0.110*	0.013	0.020	0.0219*
	0.001	0.000	0.000	0.000	0.096	0.113	0.276	0.083
Patent Citations Received Per Year	0.023	0.042	0.058	0.039	-0.737	0.006	0.342*	-0.009
	0.818	0.134	0.230	0.162	0.292	0.875	0.0621	0.903
Patent Scope	0.103	-0.0388**	-0.037	-0.0456**	-0.16	-0.0952***	-0.007	-0.031
	0.263	0.045	0.234	0.019	0.405	0.005	0.919	0.515
Lead Inventor Experience	0.118***	0.0349***	0.0380**	0.011	-0.293***	0.000	0.115	0.030
	0.002	0.000	0.011	0.265	0.00264	0.978	0.144	0.348
Patent Originality	0.318	0.208	0.309	0.560***	2.399	0.0381	-0.976	-0.107
	0.582	0.261	0.298	0.009	0.186	0.909	0.189	0.833
Device	-1.047**	-0.824***	-0.844***	-0.972***	-12.59	-0.125	-34.3	-0.523
	0.018	0.000	0.000	0.000	0.977	0.58	1.000	0.215
IPCs in Section A Only	0.425	0.234	0.289	0.409*	1.166	0.370	-0.099	-0.273
	0.558	0.305	0.399	0.093	0.423	0.325	0.898	0.554
IPCs in Section C Only	0.398	-0.048	0.024	0.161	0.208	-0.040	-0.096	-1.226**
	0.615	0.851	0.949	0.553	0.891	0.925	0.916	0.031
IPCs in Section G Only	-0.250	-0.620**	-0.457	-0.565*	-12.69	-0.287	-0.550	-0.826
	0.761	0.035	0.271	0.059	0.988	0.521	0.595	0.164
IPCs in Sections A and C Only	0.404	0.152	0.166	0.183	-0.617	-0.060	-0.151	-1.309***
	0.576	0.507	0.626	0.449	0.678	0.875	0.841	0.006
IPCs in Sections A and G Only	1.318	0.795***	0.826*	0.680**	-24.07	0.397	-0.445	-0.565
	0.207	0.005	0.061	0.019	1	0.424	0.729	0.404
IPCs in Sections A, C and G Only	1.827	0.843***	0.946**	0.779***	-0.836	0.117	-0.566	-0.6
	0.147	0.002	0.031	0.006	0.633	0.817	0.643	0.407
Patent Priority Date in 1981-85	-0.901	0.646***	0.305	0.749***	3.522	0.346	-1.815	0.614
	0.245	0.001	0.394	0.001	0.111	0.414	0.285	0.378
Patent Priority Date in 1986-90	-0.481	0.062	-0.132	-0.006	1.798	-0.286	-3.095**	-1.127**
	0.389	0.743	0.633	0.976	0.376	0.361	0.041	0.029
Patent Priority Date in 1991-95	0.163	-0.085	-0.139	-0.165	1.193	-0.711**	-2.789**	-1.177**
	0.765	0.642	0.601	0.390	0.565	0.023	0.050	0.021
Patent Priority Date in 1996-00	-0.001	0.413**	0.415	0.461**	2.454	0.307	0.105	0.177
	0.998	0.016	0.109	0.010	0.216	0.298	0.924	0.708
Constant	-0.923	-0.025	-0.132	0.109	-2.673			
	0.380	0.939	0.785	0.757	0.323			
lnalpha			-0.457***					
			0.006					
Observations	285	285	285	285	285	200	200	200
Groups								86

*** p<0.01, ** p<0.05, * p<0.1

Table 1.4b This table contains models of licensing based on patent level dependent characteristics. P-values under coefficients.

	LOGIT	**Poisson**	**Negative Binomial Regression**	**Zero Inflated Poisson**		**Cox Hazard Model**	**Cox Hazard Model** Stratified by Lead Inventor	**Cox Hazard Model** Frailty by Lead Inventor
VARIABLES	Licensed licensed=1	Times licensed	Times licensed	Times licensed	Inflate licensed=0	Time to License	Time to License	Time to License
Number of Patents Cited	0.0926***	0.0307***	0.0305***	0.008	-0.286***	0.010	0.036	0.006
	0.000	0.000	0.000	0.224	0.001	0.280	0.140	0.694
Patent Citations Received Per Year	0.048	0.0695**	0.0819*	0.038	-2.630**	-0.027	0.339	-0.060
	0.650	0.031	0.066	0.288	0.012	0.567	0.125	0.394
Patent Scope	0.143	0.005	0.007	-0.020	-0.397	-0.021	0.031	0.029
	0.111	0.827	0.823	0.435	0.164	0.567	0.685	0.566
Lead Inventor Experience	0.103***	0.0330**	0.0304*	-0.008	-0.193**	-0.0374*	0.269**	-0.045
	0.003	0.015	0.076	0.596	0.036	0.083	0.040	0.200
Patent Originality	0.384	0.389	0.417	0.267	-1.486	-0.101	0.299	-0.617
	0.527	0.150	0.215	0.441	0.297	0.830	0.712	0.266
Device	-1.014**	-1.013***	-0.973***	-1.181***	-0.444	-0.210	-39.300	-0.775*
	0.029	0.000	0.000	0.000	0.684	0.454	1.000	0.082
IPCs in Section A Only	2.641***	1.310**	1.331**	1.022*	-0.841	-2.254***	0.267	-1.915*
	0.004	0.011	0.016	0.067	0.695	0.005	0.830	0.063
IPCs in Section C Only	3.429***	1.375***	1.425**	0.710	-4.456*	-2.894***	0.676	-2.987***
	0.001	0.010	0.014	0.225	0.084	0.000	0.623	0.006
IPCs in Section G Only	1.869*	0.619	0.653	0.570	-0.271	-2.799***	0.602	-2.048*
	0.075	0.291	0.304	0.368	0.914	0.002	0.720	0.079
IPCs in Sections A and C Only	3.083***	1.248**	1.321**	0.566	-5.361**	-3.264***	-0.390	-3.839***
	0.001	0.015	0.017	0.312	0.024	0.000	0.745	0.000
IPCs in Sections A and G Only	3.711***	2.097***	2.202***	1.097*	-32.300	-2.662***	0.631	-3.023***
	0.001	0.000	0.000	0.067	1.000	0.002	0.712	0.008
IPCs in Sections A, C and G Only	3.355***	1.216**	1.234*	0.556	-4.815**	-3.808***	-0.044	-2.919**
	0.004	0.039	0.058	0.379	0.049	0.000	0.978	0.011
Patent Priority Date in 1981-85	1.438*	1.873***	1.787***	1.644***	1.010	-0.316	-19.160	-0.946
	0.080	0.000	0.000	0.001	0.689	0.542		0.253
Patent Priority Date in 1986-90	1.743***	1.471***	1.436***	1.191**	-0.864	-0.212	-19.67***	-1.096
	0.005	0.000	0.001	0.014	0.692	0.624	0.000	0.108
Patent Priority Date in 1991-95	2.239***	1.518***	1.559***	1.136**	-2.809	-0.831*	-18.09***	-0.898
	0.000	0.000	0.000	0.017	0.196	0.064	0.000	0.183
Patent Priority Date in 1996-00	0.401	1.110***	1.092**	1.371***	2.358	0.077	-20.93***	-0.617
	0.493	0.006	0.014	0.004	0.247	0.862	0.000	0.365
Constant	-5.715***	-3.190***	-3.251***	-1.472*	7.626**			
	0.000	0.000	0.000	0.073	0.033			
lnalpha					-0.893***			
					0.002			
Observations	285	285	285	285	285	152	152	152
Groups								62

*** p<0.01, ** p<0.05, * p<0.1

developments (Chatterji et al 2008). It could, however be due to the fact that new devices do not always require AMC resources or government funding to develop and the best ones may be patented outside of the AMC technology commercialization process. Alternatively, the successful ones could be developed in close collaboration with industry under sponsored research agreements and may thus be excluded from our dataset.

The results are repeated in the next two models – the Poisson and negative binomial- where the dependent variable is the number of times a patent has been considered for licensing or licensed. Because from Table 1.1a we see that both dependent variables "times looked" at and "times licensed" are over-dispersed with the standard error slightly higher than the mean, we conduct a likelihood ratio test which shows that the negative binomial, rather than the Poisson, is the appropriate model.

We also run a zero-inflated version of these models because we have an excess number of zeros in both variables (number of times looked and number of times licensed). In our sample of 285 patents, 85 patents have never been looked at and 132 have never been licensed (includes not looked at and looked at but not licensed). This zero-inflated models have two parts –a Poisson model and a logit model. The dependent variable in the Poisson part of the model is the number of times a patent has been looked at or licensed, conditional on being looked at or licensed, respectively. The separate inflation model which is a simple logit explains the excess zeros. Even though the zero inflated negative binomial model would be more appropriate, it doesn't converge and we report results from the zero inflated Poisson. A Vuong test shows that the zero inflated model is more appropriate than the regular Poisson.

The results from the zero-inflated Poisson models are similar to the previous three models. Interestingly, once controlled for the number of patents cited, a higher patent originality

predicts more agreements, conditional on there being at least one agreement on the patent. The lack of lead inventor experience seems to determine whether a patent is never looked at or licensed but not the number of times it has been looked at or licensed. Our patent scope results contradict previous literature and seem to influence our dependent variables negatively and in some models at a statistically significant level (Lerner, 1994; Decheneaux, 2009; Gambardella et al. 2007)

The last three models are Cox hazard models. Here the dependent variable is time to first agreement. Positive results indicate that as the independent variable increases, so does the hazard of an agreement. Note, however, that for ease of interpretation beta coefficients are reported rather than hazard ratios. Hazard ratios can be calculated by raising *e* to the power of the reported coefficient. We report a regular Cox hazard model in the first column, then stratified by lead inventor and then by a shared frailty (the equivalent of a random effects model) where each group is identified by a lead inventor and includes said lead inventor's patents. Our results are similar to those from the previous models but are not statistically significant in the same manner.

The models in Table 1.4b are the same as the ones in Table 1.4a, except that the dependent variables are related to licensing – i.e. licensed, times licensed, and time to license. The direction and the significance of the results are practically the same as well, except for the scope variable which is no longer negative and statistically significant. Another difference is that the "patent cites per year" variable is now positive and statistically significant in two of our models.

1.4.2 Patent-Agreement Level Models

Our patent-agreement level analyses are our main results. They test our hypothesis that technology proximity between a focal patent and a potential licensee's patent portfolio is a

determinant of whether a license will take place. All our models are logistic regressions (logits) with a dependent variable – "deal" – that is equal to one if a license was signed and zero if a deal was not done. Our full model is the last one in the respective table and it includes all our control variables. We start with a logit of our dependent variable with only the respective proximity/similarity measures. We then try a bare-bones proximity and licensee variable model. For our remaining models we start again with the fit measure and add patent citation based measures such as forward cites, backward cites and scope. We then add patent variables constructed based on cites and IPC codes - originality measure, generality measure. Technology type controls are added next – i.e. a device indicator variable and mutually exclusive dummy variables based on IPC classification by sections. Age variables are included next - technology age at time of agreement and cohort dummies for each 5 year period since 1980 based on the patent priority date. Finally, we add licensee variables – the number of granted patents that the licensee had filed before the time of the agreement and the square of the number of such patents.

Our main models in Table 1.5a test our hypotheses above that firms are more likely to license inventions that are similar but not too similar. We operationalize our technology proximity measures using the cosine variables described above. This model includes our entire sample with 588 patent–agreement level observations. We replace the cosine measures with 0s in cases in which they are not defined because the agreement firm has no patents. In table 1.5b we exclude those observations with undefined cosine measures and are left with 424 observations. As seen from the tables, the sign and statistical significance of our proximity coefficients is unchanged.

We see in these models that a higher technological proximity between the firm and the AMC patent, measured at the IPC code subclass level, is more likely to be associated with a

Table 1.5a This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. Patent level measures correspond to the hospital patent which is under the agreement. **Firms for which a cosine measure was not defined because the firm has no patents of its own were assigned a cosine measure of 0 in these models**. P-values under coefficients. Robust Standard Errors

	5.1a	5.2a	5.3a	5.4a	5.5a	5.6a	5.7a	5.8a
Cosine Subclass Level	1.157***	1.285***	1.127***	1.977***	1.361***	1.238***	1.085**	1.276***
	0.006	0.002	0.008	0.001	0.003	0.008	0.022	0.008
Within Section Cosine, Group Level	-0.921**	-1.056**	-0.968**	-2.079***	-1.259***	-1.031**	-0.856*	-1.042**
	0.032	0.014	0.028	0.001	0.010	0.037	0.094	0.041
Patent Scope			0.050	0.051	0.112***	0.105***	0.271***	0.260***
			0.122	0.229	0.005	0.010	0.000	0.000
Number of Patents Cited			0.001	0.009	0.006	0.004	-0.004	-0.006
			0.948	0.375	0.526	0.637	0.763	0.643
Patent Citations Received Per Year			0.373***	0.218**	0.404***	0.378***	0.264**	0.272**
			0.001	0.043	0.000	0.001	0.037	0.033
Patent Originality				-0.195	1.014**	0.853	1.310**	1.291**
				0.717	0.047	0.101	0.028	0.032
Device					-0.332	-0.308	-0.180	-0.110
					0.274	0.321	0.612	0.765
IPCs in Section A Only					1.539***	1.492**	1.746***	1.570**
					0.009	0.012	0.009	0.018
IPCs in Section C Only					2.551***	2.454***	3.264***	3.199***
					0.000	0.000	0.000	0.000
IPCs in Section G Only					0.768	0.828	1.588*	1.519*
					0.344	0.286	0.055	0.076
IPCs in Sections A and C Only					1.573***	1.408**	2.133***	1.923***
					0.008	0.017	0.001	0.004
IPCs in Sections A and G Only					1.948***	1.737**	2.096**	1.889**
					0.005	0.013	0.013	0.020
IPCs in Sections A, C and G Only					0.266	-0.134	-0.552	-0.605
					0.701	0.852	0.510	0.464
Technology Age in Years						0.0760***	-0.005	-0.006
						0.001	0.837	0.812
Patent Priority Date in 1981-85							1.888***	2.038***
							0.004	0.002
Patent Priority Date in 1986-90							2.088***	2.122***
							0.000	0.000
Patent Priority Date in 1991-95							2.762***	2.793***
							0.000	0.000
Patent Priority Date in 1996-00							-0.075	-0.063
							0.884	0.906
Lead Inventor Experience	-0.022	-0.015	-0.019	-0.033	-0.017	-0.019	-0.037	-0.027
	0.204	0.413	0.297	0.132	0.398	0.340	0.120	0.263
Number of Firm Patents		0.000156*						0.000265**
		0.065						0.013
Number of Firm Patents Squared		0.000						-1.78e-08**
		0.557						0.030
Patent Generality				-0.354				
				0.428				
Constant	0.110	-0.131	-0.325	0.347	-2.805***	-2.965***	-4.666***	-4.709***
	0.526	0.484	0.170	0.586	0.000	0.000	0.000	0.000
Observations	588	588	588	440	588	588	588	588

*** p<0.01, ** p<0.05, * p<0.1

Table 1.5b This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. Patent level measures correspond to the hospital patent which is under the agreement. **Firms for which a cosine measure was not defined because the firm has no patents are excluded from these models.** P-values under coefficients. Robust Standard Errors

	5.1b	5.2b	5.3b	5.4b	5.5b	5.6b	5.7b	5.8b
Cosine Subclass Level	1.039**	1.430***	0.928**	1.905***	1.434***	1.149**	1.234**	1.683***
	0.019	0.001	0.040	0.003	0.009	0.040	0.034	0.006
Within Section Cosine, Group Level	-0.983**	-1.033**	-1.046**	-2.287***	-1.508***	-1.193**	-1.211**	-1.375**
	0.023	0.017	0.019	0.000	0.006	0.034	0.042	0.022
Patent Scope			0.0694*	0.054	0.108**	0.0879*	0.324***	0.308***
			0.057	0.293	0.019	0.063	0.000	0.000
Number of Patents Cited			-0.003	0.009	0.012	0.008	-0.008	-0.010
			0.778	0.436	0.282	0.486	0.599	0.525
Patent Citations Received Per Year			0.440***	0.387**	0.541***	0.492***	0.356**	0.383**
			0.001	0.016	0.000	0.000	0.025	0.019
Patent Originality				0.729	1.492**	1.199*	1.755**	1.902**
				0.327	0.018	0.064	0.024	0.019
Device					-0.631	-0.605	-0.343	-0.235
					0.100	0.121	0.408	0.583
IPCs in Section A Only					1.773**	1.682**	2.107**	1.953**
					0.037	0.045	0.024	0.042
IPCs in Section C Only					2.723***	2.596***	3.492***	3.495***
					0.003	0.004	0.000	0.001
IPCs in Section G Only					0.364	0.555	1.529	1.358
					0.783	0.645	0.198	0.288
IPCs in Sections A and C Only					1.854**	1.673**	2.591***	2.374**
					0.031	0.048	0.007	0.016
IPCs in Sections A and G Only					2.922***	2.746***	3.320***	3.135***
					0.004	0.006	0.006	0.009
IPCs in Sections A, C and G Only					0.446	0.102	-0.730	-0.746
					0.653	0.922	0.564	0.562
Technology Age in Years						0.104***	-0.010	-0.015
						0.000	0.778	0.687
Patent Priority Date in 1981-85							1.721**	1.847**
							0.018	0.014
Patent Priority Date in 1986-90							1.819***	1.708**
							0.006	0.011
Patent Priority Date in 1991-95							2.170***	2.148***
							0.000	0.000
Patent Priority Date in 1996-00							-0.747	-0.792
							0.190	0.179
Lead Inventor Experience	-0.016	-0.008	-0.006	-0.009	-0.008	-0.005	-0.016	-0.005
	0.432	0.727	0.785	0.754	0.748	0.855	0.607	0.881
Number of Firm Patents		0.000184**						0.000268**
		0.039						0.019
Number of Firm Patents Squared		0.000						-1.70e-08**
		0.418						0.047
Patent Generality				-0.874*				
				0.0913				
Constant	0.168	-0.304	-0.337	-0.233	-3.500***	-3.598***	-5.053***	-5.342***
	0.435	0.226	0.250	0.766	0.001	0.001	0.000	0.000
Observations	424	424	424	304	424	424	424	424

*** p<0.01, ** p<0.05, * p<0.1

license i.e. a "deal." Holding subclass level proximity constant, however, a higher technological proximity measured at the main group level of the IPC codes is less likely to result in a deal.

This result is hard to comprehend because of the complexity of the measures. A stylized example may be useful to let us abstract from the details of the patent classes by using more easily understandable consumer product categories that may be protected by these patents. Note that this is a simplification for clarification purposes– our IPC subclass level and IPC main group levels do not necessarily correspond to the product categories described here, including their breath and generality.

Imagine a medical device company A that has 10 patents - 5 for metal and 5 for ceramic prosthetic hip implants. These technologies are all in the 'prosthetic hips' area at the broad level of measurement but different at the more granular level – 'type of prosthetic hip.' Another firm B also has 10 patents but all 10 for plastic hip implants. A third firm C, also a medical device company has patents for stents and catheters but none in the hip implants area. All three firms have signed confidentiality agreements showing interest in an AMC patent of a *plastic* hip implant. If we only looked at the broad level measures, both firm A and firm B would be more likely to license the new plastic hip implant patent that is offered by our AMC than firm C – their technology proximity measure to the patent offered is higher than that of firm C. Also, firm A would be as equally likely to license it as firm B as their proximity measure values are the same at this level.

Now let's look at firm A and firm B, at the more granular level of proximity, controlling for the higher level proximity. Based on our results we would expect that firm A is more likely to license the AMC patent than firm B. While firm B already has plastic hip implant patents, firm A

has none, so at the granular level of measurement, firm B's portfolio is closer to the focal patent than firm A's portfolio which makes it less likely to license it.

These results are in fact intuitive. A match at the broad level (i.e. hip implants) is useful since it indicates that both firms may have specific complementary technologies to the focal patent. One example would be that both firms have their own technology that deals with specific implant shape (i.e. implant is hollow inside to enhance movement) or components needed to attach the implant. These broad level technologies would be likely relevant to implants of any materials. In general, their existing broad level complementary technologies will increase their marginal benefit of licensing any patent for which they are relevant. However, firm B already has its own *plastic* implant technology. If it licensed our focal patent, it would most likely be duplicating its own technology at a cost. Furthermore, choosing outside *plastic* hip technology over internally developed one may be harder given the lack of knowledge about the outside technology and the difficulty in transferring tacit knowledge (Polanyi, 1966; von Hippel, 1988; Agarwal, 2006). This implies that a higher similarity measured at the more granular, detailed level is likely to not result in a license.

To understand whether these results are different for different types of firms, we split our sample by firm R&D age and by firm size. Our goal is to separate firms that are based on different types of technology platforms - young biotech firms that do mostly large molecule drug development and older pharmaceutical firms specializing in small molecule drug discovery. Our samples are not entirely "clean" in the sense that there are also medical device and software companies in both groups. Furthermore, the distinction between biotech and pharmaceutical firms has grown blurrier over the years as pharmaceutical firms have also developed large molecule drug technologies and young biotechnology companies who started based on one

specific invention have developed related expertise. Interaction effects would have been better for these exercises but they would have complicated interpretation significantly because we have two different variables we want to interact with multiple size and age variables.

In Tables 1.6a through 1.6c we split our sample in different groups by age - firms with an R&D age of less than 10 years, those with an R&D age between 10 and 20 years and firms older than 20 years. We note that our results have the same sign as in the full sample models but are not statistically significant in many of the models in the first age group, especially the measures based on more granular IPC code slices at the IPC main group level. Our results are, however very strong in terms of both size and statistical significance in the 10-20 year old firms group. They are also strong in a few of the models in the older firm group. Note however, that the addition of technology age and technology cohort variables changes the value and significance of the proximity measures indicating that they are related. It is also interesting to point out here that technology age is negative and significant among medium aged firms (10-20 years of R&D) and positive and significant for older firms (20+ years of R&D). This implies that older firms are more likely to license older and more established technologies, while younger firms may be more willing to undertake risks. This result is not driven by inventor startups, however, as those would have no patents at the time of licensing and would not be in any of the age groups. Note that several of our control variables for IPC code section and cohort drop resulting in smaller sample sizes for some of the models.

We then go on to split our samples by size. By looking at firm names and the number of patents that they own at time of licensing we find out that most biotechnology firms have fewer than 500 patents and most big pharmaceutical companies have more than 500 patents. Again, this division is not “clean” in the sense discussed above. The sign of our major results hold again.

Table 1.6a Sample of firms that are older than 0 and younger than 10 years at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm technology age (i.e. time between the first patent that the firm filed and the agreement that it signed for the AMC patent) is smaller than 10 years. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients. Robust Standard Errors

	6.1a	6.2a	6.3a	6.4a	6.5a	6.6a	6.7a	6.8a
Cosine Subclass Level	0.763	0.818	0.959	1.529	2.012*	2.007*	2.919*	2.934*
	0.333	0.315	0.232	0.239	0.096	0.098	0.051	0.079
Within Section Cosine, Group	0.455	0.533	-0.158	-0.776	-1.631	-1.614	-3.708*	-3.305
	0.584	0.528	0.861	0.552	0.242	0.259	0.075	0.108
Patent Scope			-0.079	-0.159	-0.197	-0.198	-0.190	-0.200
			0.603	0.472	0.453	0.456	0.502	0.536
Number of Patents Cited			-0.029	-0.018	-0.036	-0.035	-0.048	-0.046
			0.236	0.517	0.311	0.343	0.341	0.368
Patent Cites Received Per Year			0.686**	0.525	1.143***	1.132***	1.211**	1.469**
			0.026	0.114	0.002	0.003	0.027	0.022
Patent Originality				0.459	0.634	0.615	-0.859	-0.627
				0.723	0.647	0.662	0.598	0.736
Device					1.327	1.311	1.660	1.009
					0.131	0.140	0.225	0.474
IPCs in Section A Only					15.49***	16.36***	18.64***	19.40***
					0.000	0.000	0.000	0.000
IPCs in Section C Only					15.67***	16.54***	20.15***	21.22***
					0.000	0.000	0.000	0.000
IPCs in Section G Only					13.95***	14.84***	17.67***	18.13***
					0.000	0.000	0.000	0.000
IPCs in Sections A and C Only					16.76***	17.63***	21.89***	22.24***
					0.000	0.000	0.000	0.000
IPCs in Sections A and G Only					15.17***	16.03***	0.006	1.687
					0.000	0.000	0.998	0.578
Lead Inventor Experience	-0.004	0.005	0.010	0.018	0.002	0.002	0.000	0.011
	0.919	0.904	0.845	0.773	0.976	0.980	0.998	0.863
Number of Firm Patents		0.062						0.107*
		0.103						0.091
Number of Firm Patents Squared		-0.001						-0.001
		0.299						0.183
Patent Generality				-1.616				
				0.236				
Technology Age in Years						0.006	-0.097	-0.085
						0.934	0.319	0.462
Patent Priority Date in 1986-90							18.24***	17.63***
							0.000	0.000
Patent Priority Date in 1991-95							1.112	0.963
							0.366	0.481
Patent Priority Date in 1996-00							-2.200**	-2.520**
							0.037	0.019
Constant	-1.077**	-1.661***	-1.114	-0.173	-17.47***	-18.34***	-19.05***	-20.76***
	0.042	0.007	0.137	0.906	0.000	0.000	0.000	0.000
Observations	110	110	110	76	105	105	98	98

*** p<0.01, ** p<0.05, * p<0.1

Table 1.6b: Sample of firms that are older than 10 and younger than 20 years at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm technology age (i.e. time between the first patent that the firm filed and the agreement that it signed for the AMC patent) is larger than 10 and smaller than 20. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients. Robust Standard Errors

	6.1b	6.2b	6.3b	6.4b	6.5b	6.6b	6.7b	6.8b
Cosine Subclass Level	1.610**	1.787**	1.930**	2.634***	2.172**	3.022***	4.013***	4.397**
	0.021	0.046	0.013	0.002	0.010	0.002	0.003	0.027
Within Section Cosine, Group	-2.331***	-2.456**	-2.533***	-3.507***	-3.074***	-4.145***	-3.993***	-3.781***
	0.002	0.012	0.003	0.000	0.001	0.000	0.001	0.006
Patent Scope			-0.129	-0.158	0.023	0.087	0.542*	0.593
			0.333	0.364	0.901	0.624	0.092	0.118
Number of Patents Cited			0.039	0.053	0.031	0.024	0.0935*	0.111*
			0.159	0.138	0.425	0.585	0.056	0.075
Patent Cites Received Per Year			0.196	0.139	0.425	0.472*	0.404	0.482
			0.501	0.710	0.137	0.095	0.205	0.174
Patent Originality				-1.115	0.691	1.434	0.483	1.305
				0.502	0.733	0.505	0.815	0.571
Device					-0.541	-0.984	-2.839**	-2.752**
					0.459	0.210	0.017	0.046
IPCs in Section A Only					2.611**	3.501***	18.70***	17.98***
					0.032	0.005	0.000	0.000
IPCs in Section C Only					3.341***	3.901***	20.70***	20.42***
					0.010	0.003	0.000	0.000
IPCs in Sections A and C Only					1.288	1.915*	17.51***	16.99***
					0.273	0.092	0.000	0.000
IPCs in Sections A and G Only					2.730	4.103**	19.30***	18.72***
					0.108	0.021	0.000	0.000
IPCs in Sections A, C and G Only					0.958	1.817	16.74***	16.02***
					0.469	0.150	0.000	0.000
Lead Inventor Experience	0.014	0.015	-0.034	-0.052	-0.051	-0.045	-0.065	-0.100
	0.762	0.743	0.539	0.397	0.358	0.416	0.308	0.186
Number of Firm Patents		-0.001						-0.00803**
		0.581						0.020
Number of Firm Patents Squared		0.000						5.89e-06*
		0.514						0.074
Patent Generality				-1.044				
				0.310				
Technology Age in Years						-0.179***	-0.182**	-0.254**
						0.00484	0.042	0.024
Patent Priority Date in 1981-85							-1.687	-0.577
							0.395	0.832
Patent Priority Date in 1986-90							-1.366	-1.210
							0.392	0.534
Patent Priority Date in 1996-00							-3.448**	-3.653*
							0.0245	0.063
Constant	0.432	0.461	0.558	2.174	-2.183	-2.668	-17.56***	-17.02***
	0.342	0.349	0.434	0.276	0.386	0.297	0	0.000
Observations	134	134	134	108	133	133	114	114

*** p<0.01, ** p<0.05, * p<0.1

Table 1.6c: Sample of firms that are older than 20 years at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation is a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm technology age (i.e. time between the first patent that the firm filed and the agreement that it signed for the AMC patent) is larger than 20 years. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients. Robust Standard Errors

	6.1c	6.2c	6.3c	6.4c	6.5c	6.6c	6.7c	6.8c
Cosine Subclass Level	2.116**	2.827***	1.998*	2.893*	3.256***	0.802	0.747	2.926*
	0.024	0.002	0.055	0.060	0.007	0.538	0.531	0.078
Within Section Cosine, Group	-1.626**	-1.336**	-1.571**	-2.974**	-2.970***	-1.626	-1.238	-1.678
	0.018	0.045	0.040	0.018	0.004	0.180	0.373	0.253
Patent Scope			0.142**	0.173*	0.123	0.142	0.300**	0.467**
			0.011	0.053	0.107	0.176	0.042	0.018
Number of Patents Cited			-0.011	-0.006	0.0377*	0.013	-0.003	-0.006
			0.432	0.748	0.058	0.573	0.887	0.847
Patent Cites Received Per Year			0.533**	0.613**	0.970***	0.941***	0.668**	0.826**
			0.013	0.041	0.002	0.001	0.036	0.015
Patent Originality				1.759	2.738**	2.780*	2.344	2.780
				0.141	0.041	0.055	0.180	0.174
Device					-3.125***	-2.463**	-1.232	-1.058
					0.001	0.020	0.294	0.472
IPCs in Section A Only					-0.084	-0.156	-1.048	-1.640
					0.937	0.856	0.413	0.192
IPCs in Section C Only					2.171*	2.056**	1.225	1.989
					0.054	0.027	0.302	0.108
IPCs in Section G Only					3.818**	3.891**	1.352	1.069
					0.011	0.038	0.512	0.631
IPCs in Sections A and C Only					0.673	0.549	-0.181	-1.204
					0.514	0.518	0.896	0.379
IPCs in Sections A and G Only					2.360*	2.712**	1.785	1.779
					0.086	0.019	0.370	0.349
IPCs in Sections A, C and G Only					-0.364	0.167	-0.152	-0.675
					0.801	0.911	0.915	0.669
Lead Inventor Experience	-0.010	-0.029	0.003	-0.002	-0.031	0.014	-0.003	0.036
	0.780	0.441	0.938	0.960	0.543	0.799	0.956	0.588
Number of Firm Patents		0.0003**						0.00076***
		0.037						0.000
Number of Firm Patents Squared		0.000						-4.50e-08**
		0.275						0.000
Patent Generality				-1.169				
				0.177				
Technology Age in Years						0.319***	0.217***	0.316***
						0.000	0.007	0.001
Patent Priority Date in 1986-90							4.381**	5.029***
							0.024	0.010
Patent Priority Date in 1991-95							2.655	2.935
							0.155	0.113
Patent Priority Date in 1996-00							1.324	1.601
							0.458	0.357
Constant	0.274	-0.718	-0.558	-1.112	-3.258**	-5.041***	-5.579***	-9.334***
	0.416	0.106	0.238	0.369	0.033	0.001	0.001	0.000
Observations	180	180	180	120	180	180	180	180

*** p<0.01, ** p<0.05, * p<0.1

Table 1.7a Sample of firms that have more than 0 and less then 500 patents at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm has more than 0 and less then 500 patents. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients.

	7.1a	7.2a	7.3a	7.4a	7.5a	7.7a	7.7a	7.8a
Cosine Subclass Level	0.956*	0.915*	1.074**	2.043***	1.498**	1.466**	1.325*	1.455*
	0.080	0.091	0.048	0.008	0.028	0.034	0.081	0.075
Within Section Cosine, Group	-0.278	-0.349	-0.436	-1.598*	-0.766	-0.741	-0.801	-1.289
	0.646	0.563	0.483	0.058	0.320	0.341	0.377	0.199
Patent Scope			-0.176*	-0.167	-0.111	-0.114	0.334**	0.331**
			0.057	0.162	0.285	0.276	0.033	0.032
Number of Patents Cited			-0.014	0.005	-0.008	-0.007	-0.005	-0.008
			0.299	0.767	0.672	0.698	0.820	0.745
Patent Cites Received Per Year			0.470***	0.353*	0.585***	0.566***	0.272	0.413*
			0.008	0.089	0.000	0.001	0.158	0.059
Patent Originality				0.285	1.760**	1.683**	2.000**	1.806*
				0.746	0.019	0.027	0.028	0.070
Device					-0.386	-0.377	-0.514	-0.714
					0.333	0.343	0.301	0.175
IPCs in Section A Only					1.733*	1.707*	2.458**	2.233**
					0.058	0.062	0.015	0.037
IPCs in Section C Only					1.890*	1.892*	3.583***	3.574***
					0.057	0.055	0.002	0.002
IPCs in Section G Only					-1.329	-1.266	0.818	0.651
					0.395	0.413	0.602	0.684
IPCs in Sections A and C Only					1.149	1.138	2.556**	2.494**
					0.211	0.214	0.015	0.023
IPCs in Sections A and G Only					2.452**	2.377**	2.450*	2.987**
					0.033	0.041	0.091	0.041
IPCs in Sections A, C and G Only					-0.374	-0.425	-0.841	-0.915
					0.744	0.718	0.540	0.526
Technology Age in Years						0.018	-0.042	-0.053
						0.633	0.382	0.270
Patent Priority Date in 1981-85							1.241	1.310
							0.198	0.185
Patent Priority Date in 1986-90							1.485*	1.049
							0.065	0.189
Patent Priority Date in 1991-95							2.219***	2.398***
							0.003	0.001
Patent Priority Date in 1996-00							-1.132*	-1.310*
							0.078	0.052
Lead Inventor Experience	-0.007	-0.005	-0.016	-0.011	-0.013	-0.013	-0.003	0.002
	0.794	0.839	0.589	0.761	0.704	0.704	0.934	0.963
Number of Firm Patents		0.005						0.0147**
		0.225						0.011
Number of Firm Patents Squared		0.000						-4.67e-05***
		0.173						0.005
Patent Generality				-0.933				
				0.156				
Constant	-0.331	-0.396	0.038	0.292	-2.953**	-2.955**	-5.220***	-5.073***
	0.251	0.210	0.930	0.771	0.026	0.026	0.001	0.002
Observations	266	266	266	191	266	266	266	266

*** p<0.01, ** p<0.05, * p<0.1

Table 1.7b Sample of firms that have more than 500 patents at time of agreement

This table shows logit models with a dependent variable equal to 1 if the confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair - a patent can have multiple agreements and each agreement can be associated with multiple patents. The sample is restricted to agreements where the firm has more than 500 patents. Patent level measures correspond to the hospital patent which is under the agreement. P-values under coefficients. Robust Standard Errors

	7.1b	7.2b	7.3b	7.4b	7.5b	7.7b	7.7b	7.8b
Cosine Subclass Level	1.510	1.683	0.335	0.315	3.400*	0.493	1.803	7.380***
	0.237	0.159	0.819	0.866	0.061	0.812	0.365	0.008
Within Section Cosine, Group	-2.406***	-1.996***	-2.605***	-4.194***	-4.014***	-2.920**	-2.787**	-3.484***
	0.000	0.003	0.001	0.001	0.000	0.015	0.024	0.004
Patent Scope			0.171**	0.159*	0.056	0.057	0.257	0.475**
			0.013	0.050	0.511	0.582	0.121	0.047
Number of Patents Cited			0.006	0.017	0.0619***	0.0443*	0.016	0.034
			0.661	0.453	0.007	0.078	0.532	0.343
Patent Cites Received Per Year			0.667**	0.868*	1.367***	1.257***	1.113***	1.446***
			0.025	0.053	0.003	0.001	0.005	0.002
Patent Originality				1.469	2.614	2.615	1.548	3.103
				0.387	0.107	0.128	0.464	0.311
Device					-3.683***	-3.017**	-1.310	-1.775
					0.004	0.018	0.317	0.286
IPCs in Section A Only					13.60***	13.53***	9.123***	9.607***
					0.000	0.000	0.000	0.000
IPCs in Section C Only					16.44***	15.46***	12.29***	14.25***
					0.000	0.000	0.000	0.000
IPCs in Section G Only					18.15***	17.64***	12.49***	13.81***
					0.000	0.000	0.000	0.000
IPCs in Sections A and C Only					14.38***	14.10***	10.46***	10.48***
					0.000	0.000	0.000	0.000
IPCs in Sections A and G Only					16.85***	16.95***	13.47***	15.03***
					0.000	0.000	0.000	0.000
IPCs in Sections A, C and G Only					14.57***	14.51***	11.17***	12.00***
					0.000	0.000	0.000	0.000
Technology Age in Years						0.247***	0.260***	0.275**
						0.000	0.007	0.013
Patent Priority Date in 1981-85							4.600**	7.769***
							0.048	0.000
Patent Priority Date in 1986-90							3.913*	5.313***
							0.054	0.005
Patent Priority Date in 1991-95							1.843	2.079
							0.366	0.270
Patent Priority Date in 1996-00							1.495	1.150
							0.470	0.556
Lead Inventor Experience	-0.004	-0.009	0.022	0.017	-0.032	0.008	0.033	0.040
	0.918	0.815	0.614	0.768	0.601	0.895	0.691	0.659
Number of Firm Patents		0.000						0.00108***
		0.321						0.001
Number of Firm Patents Squared		0.000						-5.66e-08**
		0.689						0.003
Patent Generality				-1.282				
				0.198				
Constant	0.797*	0.066	-0.166	-0.210	-16.71***	-17.56***	-15.89***	-23.23***
	0.069	0.910	0.755	0.890	0.000	0.000	0.000	0.000
Observations	158	158	158	113	158	158	158	158

*** p<0.01, ** p<0.05, * p<0.1

Interestingly, the results are only statistically significant for the cosine measure at the subclass level in the smaller sized group and only for the cosine measure at the more granular, main group level for the larger firms. This implies that finding a closely related technology is more important for smaller firms while not duplicating efforts may be more important for larger firms. Note, however, that controlling for size in the sample of firms with more than 500 patents gives us very large and highly significant coefficients for both proximity measures in the full model.

The next set of models in Table 1.8 includes only the first agreement that is signed for a patent whether it is a "deal" or "no deal." We are interested in these results because we are concerned that whether the first agreement is a "deal" or "no deal" may signal patent quality and may influence future licenses, especially non-exclusive licenses of which there may be potentially many per patent. An exclusively licensed invention, on the other hand, takes the patent off the market. Of the 200 patents that have at least one agreement, 6 are excluded because of licensee issues (discussed in the data section) and we are left with 194 patent-first agreement pairs. Our technology proximity results from the previous models still hold in this sample and are statistically significant indicating that the results are robust and are not driven by a few patents that have been licensed multiple times since each patent appears only once in this dataset.

1.6 Conclusion

In this paper we addressed a gap in the literature on markets for technology by taking a close look at the demand for technology. While this has been attempted in previous papers, our unique dataset that includes not only firms that licensed technologies but also showed interest in them but did not license provides an important control group for our description of the structure of such markets. We showed that proximity matters in the technologies firms decided to license.

Table 1.8 First Agreement Models

This table shows logit models with a dependent variable equal to 1 if the FIRST confidentiality agreement became a license (i.e. "deal") and 0 if the agreement did not result in a license (i.e. "no deal"). Each observation corresponds to a patent-agreement pair.**Only the first agreement for each patent was selected.** Patent level measures correspond to the hospital patent which is under the agreement.**Firms for which a cosine measure was not defined because the firm has no patents of its own were assigned a cosine measure of 0 in these models**. P-values under coefficients. Robust Standard Errors

	8.1		8.2	8.3	8.4	8.5	8.7	8.7	8.8
Cosine Subclass Level	2.095**	1.465*	1.664*	3.993***	1.847*	1.722	2.725**	2.744**	
	0.017	0.090	0.065	0.002	0.090	0.114	0.021	0.021	
Within Section Cosine, Group	-2.651***	-1.856**	-2.325**	-4.984***	-2.507**	-2.519**	-2.473**	-2.460**	
	0.002	0.033	0.011	0.001	0.029	0.022	0.018	0.031	
Patent Scope			0.101**	-0.003	0.145*	0.106	0.438**	0.448***	
			0.042	0.971	0.051	0.252	0.011	0.009	
Number of Patents Cited			-0.001	-0.012	0.011	0.008	0.010	0.012	
			0.943	0.608	0.540	0.618	0.731	0.660	
Patent Cites Received Per Year			0.299	0.062	0.344	0.350	-0.137	-0.134	
			0.186	0.531	0.128	0.113	0.109	0.110	
Patent Originality				0.207	0.964	0.935	0.953	1.159	
				0.861	0.252	0.287	0.494	0.469	
Device					0.097	0.031	0.343	0.380	
					0.857	0.957	0.645	0.603	
IPCs in Section A Only					2.117**	2.331**	3.986**	3.803**	
					0.020	0.031	0.010	0.013	
IPCs in Section C Only					3.147***	3.164***	6.434***	6.290***	
					0.002	0.006	0.003	0.002	
IPCs in Section G Only					1.287	1.287	4.725**	4.085**	
					0.258	0.311	0.016	0.022	
IPCs in Sections A and C Only					2.074**	2.057*	4.641**	4.441**	
					0.022	0.053	0.015	0.018	
IPCs in Sections A and G Only					1.924	1.993	3.914	4.154	
					0.135	0.154	0.136	0.146	
IPCs in Sections A, C and G Only					2.584**	2.772**	4.991**	6.382*	
					0.028	0.035	0.023	0.054	
Technology Age in Years						0.161**	-0.058	-0.132	
						0.021	0.588	0.410	
Patent Priority Date in 1981-85							3.338**	3.556**	
							0.027	0.039	
Patent Priority Date in 1986-90							5.322***	5.178***	
							0.000	0.000	
Patent Priority Date in 1991-95							4.809***	4.518***	
							0.000	0.001	
Patent Priority Date in 1996-00							-0.249	-0.300	
							0.788	0.760	
Lead Inventor Experience			-0.026	-0.0989**	-0.024	-0.024	-0.065	-0.064	
			0.436	0.038	0.518	0.528	0.233	0.258	
Number of Firm Patents		-0.000661**						-0.001	
		0.010						0.306	
Number of Firm Patents Squared		6.67e-08***						0.000	
		0.004						0.272	
Patent Generality				-3.808***					
				0.003					
Constant	0.576***	0.515**	0.267	4.183**	-2.811**	-3.138**	-7.631***	-7.324**	
	0.006	0.022	0.480	0.014	0.026	0.035	0.010	0.014	
Observations	194	194	194	145	194	194	194	194	

*** p<0.01, ** p<0.05, * p<0.1

Our identification comes from variation within a group that showed at least a threshold level of interest in the technology by contacting the licensing office and signing a confidentiality agreement. Future research will expand on this by identifying a larger population of potential buyers in this market based on some other measure of interest – we currently do not include informal channels through which information may have been obtained or inquiries that did not result in signing of a confidentiality agreement.

We also contribute to the literature on measurement of technology proximity by using a new patent statistic – the international patent class which with its nested structure allows for proximity measurement at the broad as well as granular level between different (portfolios of) patents. We also improved on existing measures by including multiple classes rather than just one, resulting in more robust results (Benner and Weldfogel, 2008). Further comparison and validation of these new measures is in order.

Ultimately the real question is whether these technologies make it to the product market once they are licensed and how the technology proximity, either at the broad or the granular level influences that outcome. It would be interesting to know whether in-licensed technologies that are very close to the licensee's in-house developed technology are strategically shelved or perhaps not absorbed by the firm due to behavioral resistance to outside innovations, the so called "not-invented-here" syndrome (Katz and Allen, 1982; Thursby and Thursby, 2004). We view this paper as a first step in this exciting direction.

1.7 Bibliography

Agarwal, AK (2006), "Engaging the Inventor: Exploring Licensing Strategies for University Inventions and the Role of Latent Knowledge" *Strategic Management Journal*, 27 (1), pp. 63-79

Anand B.N. and T. Khanna (2000),"The structure of licensing contracts", Journal *of Industrial Economics*, 48 (1), pp. 103 – 135

Anton, J.J. and D.A. Yao (1994), "Expropriation and Inventions: Appropriable Rents in the Absence of Property Rights." American Economic Review, vol. 84 pp.190-209.

Arora, A., Fosfuri, A., and Gambardella, A. (2001) Markets for Technology – The Economics of Innovation and Corporate Strategy, Cambridge, MA: MIT Press

Arora, A. and A. Gambardella (1994), 'Evaluating technological information and utilizing it: scientific knowledge, technological capability, and external linkages in biotechnology,' *Journal of Economic Behavior & Organization*, 24(1), pp. 91–114

Arora, A. and A. Gambardella (2010), "Ideas for Rent: An Overview of Markets for Technology," *Industrial and Corporate Change*, vol. 19 (3) pp. 775-803

Arora, A. and A. Fosfuri (2003), 'Licensing the market for technology,' *Journal of Economic Behavior and Organization*, 52(2), pp. 277–295.

Arrow, K. (1962), "Economic Welfare and the Allocation of Resources for Invention," in: *The Rate and Direction of Inventive Activity: Economic and Social Factors*, National Bureau of Economic Research, Inc., pp.609-626

Athreye, S. and J. Cantwell (2007), 'Creating competition? Globalization and the emergence of new technology producers,' *Research Policy*, 36(2), pp. 209–226

Atuahenegima, K (1993), "Buying Technology for Product Development in Smaller Firms", *Industrial Marketing Management*, 22 (3), pp. 223-232

Atuahenegima, K. (1992), Inward Technology Licensing As an Alternative to Internal Research-and-Development in New Product Development - a Conceptual-Framework, *Journal of Product Innovation Management*, 9(2), pp.156-167

Atuahenegima, K. and P. Patterson (1993), "Managerial Perceptions of Technology Licensing as an Alternative to Internal Research-and-Development in New Product Development - an Empirical-Investigation," *R & D Management*, 23 (4), pp. 327-336

Benkard, L. (2000), "Learning and Forgetting: The Dynamics of Aircraft Production", *American Economic Review,* 90 (4), pp. 1034 – 1054

Benner, M. and J. Waldfogel (2008), "Close to you? Bias and precision in patent-based measures of technological proximity", *Research Policy,* 37(9), pp. 1556-1567

Bessy, C and E. Brousseau, (1998) "Technology licensing contracts features and diversity", *International Review of Law and Economics*, 18(4) pp. 451-489

Breschi S., F. Lissoni F and F. Malerba (2003), "Knowledge-relatedness in firm technological diversification," *Research Policy,* 32 (1), pp. 69-87

Bresnahan, T. and M. Trajtenberg (1995), "General purpose technologies 'Engines of growth'?" *Journal of Econometrics*, 65(1), pp. 83-108

Cassiman B. and R. Vergeulers (2006), In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition, *Management Science*, 52 (1), pp. 68-82

Caves, R., H. Crookell and JP Killing (1983), "The Imperfect Market for Technology Licenses," *Oxford Bulletin of Economics and Statistics*, 45 (3), pp.246-267

Chatterji, A.K., K. Fabrizio, W. Mitchell and K.A. Schulman (2008), Physician-Industry Cooperation in the Medical Device Industry, *Health Affairs*, 27(6), pp. 1532-1543

Cockburn, I.M., M.J. MacGarvie and E. Muller (2010) Patent thickets, licensing and innovative performance, *Industrial And Corporate Change*, 19 (3), pp. 899-925

Cohen, W. and D. Levinthal (1990), "Absorptive Capacity: A New Perspective on Learning and Innovation". *Administrative Science Quarterly*, 35(1), pp. 128-152

Dechenaux E., B. Goldfarb, S. Shane and M. Thursby (2008), Appropriability and Commercialization: Evidence from MIT Inventions, *Management Science*, 54 (5), pp. 893-906

Ford, D. (1988), "Develop Your Technology Strategy," Long Range Planning, 11(5), pp. 85 -95

Gale, D. and L. S. Shapley (1962), "College Admissions and the Stability of Marriage", *American Mathematical Monthly*, 69 (1), pp.9-14

Gambardella, A., P. Giuri and A. Luzzi (2007), "The market for patents in Europe," Research Policy, 36(8), pp.1163–1183

Gans, J., D. H. Hsu and S. Stern (2008), "The impact of uncertain intellectual property rights on the market for ideas: evidence from patent grant delays," Management Science, 54(5), pp.982–997.

Gans, J. and S. Stern (2010), "Is There a Market for Ideas", *Industrial and Corporate Change*, 19(3), pp. 805-837

Haagedorn, J., Lorenz-Orlean S. and H. Kranenburg (2009), "Technology Transfer: Partnership Embedded Licensing Agreements", Industrial and Corporate Change, 18 (3), pp. 529-550

Hall, B., A. Jaffe and M. Trajtenberg (2001), "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper No. 8498

Heller, MA and RS Eisenberg (1998), Can Patents Deter Innovation? The Anticommons in Biomedical Research, *Science, v*ol. 280 (5364), pp. 698-701

Henderson, R., A. Jaffe & M. Trajtenberg (1998), "Universities as a Source of Commercial Technology*:* A Detailed Analysis of University Patenting, 1965-1988", *Review of Economics and Statistics,* Vol. 80(1), pp. 119-127

Higgins, M. J. and D. Rodriguez (2006), "The outsourcing of R&D through acquisitions in the pharmaceutical industry," *Journal of Financial Economics,* 80(2), pp. 351–383

Jaffe, Adam B. (1986) "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value," American Economic Review, 76, pp. 984–1001

Katz, R. and T. Allen (1982), Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R & D Project Groups, *R&D Management*, 12 (1), pp. 7-20

Killing, JP (1978), "Diversification through Licensing", R&D Management, 8(3), pp. 159 – 163

Kurokawa S. (1997), "Make-or-buy decisions in R&D: Small technology based firms in the United States and Japan," *IEEE Transactions on Engineering Management,* 44 (2), pp. 124-134

Laursen, K; M, Leone, and S. Torrisi, (2010), "Technological exploration through licensing: new insights from the licensee's point of view," *Industrial and Corporate Change*, 19 (3) pp. 871 - 897

Lerner, J (1994), The Importance Of Patent Scope - An Empirical-Analysis , *Rand Journal Of Economics*, 25 (2), pp. 319-333

Lichtenthaler, U. (2007), "The drivers of technology licensing: an industry comparison," California Management Review, 49(4), pp. 67–89.

Lichtenthaler, U. and H. Ernst (2007), "External technology commercialization in large firms: results of a quantitative benchmarking study," R&D Management, 37(5), pp. 383–397

Lowe J. and P. Taylor (1998) R & D and Technology Purchase through License Agreements: Complementary Strategies and Complementary Assets, *R & D Management*, 28 (4), pp. 263-278

Mehta,A., M. Rysman and T. Simcoe (2010),"Identifying the Age Profile of Patent Citations: New Estimates of Knowledge Diffusion", *Journal of Applied Econometrics*, 25 (7), pp. 1179-1204

Mowery D., J. Oxley and B. Silverman (1996), "Strategic Alliances And Interfirm Knowledge Transfer," *Strategic Management Journal*, 17 (SI), pp.77-91

Narin F., E. Noma and R. Perry (1987) "Patents As Indicators of Corporate Technological Strength ," *Research Policy* , 16 (2-4), pp. 143-155

Nerkar A. and S. Shane (2007), "Determinants of invention commercialization: An empirical examination of academically sourced inventions," *Strategic Management Journal*, 28 (11), pp.1155-1166

Patel, P and K. Pavitt (1994), "The Continuing, Widespread (and Neglected) Importance of Improvements In Mechanical Technologies ," *Research Policy*, 23, pp. 533-545

Pisano, G. (1990) "The R&D Boundaries of the Firm: An Empirical Analysis," *Administrative Science Quarterly*, 35 (1) pp. 153-176

Pitkethly, R.H. (2001), "Intellectual property strategy in Japanese and UK companies: patent licensing decisions and learning opportunities", *Research Policy*, 30(3), pp. 425–442

Polanyi, M (1966), The Tacit Dimension, Garden City, N.Y. : Doubleday

Razgaitis, R. (2004), "US/Canadian licensing in 2003: survey results," *Journal of the Licensing Executive Society*, 34(4), pp.139–151

Robbins, C. (2006), 'Measuring payments for the supply and use of intellectual property,' Bureau of Economic Analysis, U.S. Department of Commerce: Washington, DC

Rosenberg, N. (1990). "Why Do Firms Do Basic Research (With Their Own Money)?" *Research Policy, 19*(2), pp. 165-174.

Roth, A. E. (2008) "What have we learned from market design?" Hahn Lecture, *Economic Journal*, 118 (March), pp. 285–310

Sheehan, J., C. Martinez and D. Guellec (2004), 'Understanding business patenting and licensing: results of a survey,' OECD: Paris.

Silverman B. (1999), “Technological Resources and the Direction of Corporate Diversification: Toward an Integration of the Resource-Based View and Transaction Cost Economics,” *Management Science*, 46 (8), pp. 1109-1124

Teece, D. (1986), “Profiting from Technological Innovation: Implications for integration, collaboration, licensing and public policy”, *Research Policy*, 15 (6), pp. 285-305

Thursby J.G. and M.C. Thursby (2004), “Are faculty critical? Their role in university-industry licensing”, *Contemporary Economic Policy*, 22 (2), pp. 162-178

Trajtenberg, M. (1990) A Penny for your Quotes - Patent Citations and the Value of Innovations, *RAND Journal of Economics*, 21 (1), pp. 172-187

Tsai, K. and J.Wang, (2007), Inward technology licensing and firm performance: a longitudinal study, *R & D Management*, 37 (2), pp. 151-160

Tsai, K. and J.Wang, (2009) “External technology sourcing and innovation performance in LMT sectors: An analysis based on the Taiwanese Technological Innovation Survey”, *Research Policy*, 38 (3), pp.518–526

von Hippel, E. (1976) "The Dominant Role of Users in the Scientific Instrument Innovation Process," *Research Policy*, 5(3), pp. 212-39

von Hippel, E. (1988), The Sources of Innovation , New York: Oxford University Press

Zeckhauser R. (1996), “The Challenge of Contracting for Technological Information," *Proceedings of the National Academy of Sciences*, 93, pp.12743-12748

Zuniga, M. P. and D. Guellec (2008), ‘Survey on patent licensing: initial results from Europe and Japan,’ OECD: Paris.

2. TRANSLATING INVENTIONS INTO PRODUCTS: INVENTORS' EDUCATIONAL BACKGROUND AND TECHNOLOGY LICENSING FROM ACADEMIC MEDICAL CENTERS

2.1 Abstract

The great leaps that have been made in basic life sciences in recent decades have brought to light the need for translating these research findings into practical applications.[1] NIH's support for *translational research* emphasizes the need for interdisciplinary and multidisciplinary team collaboration as well as training of interdisciplinary researchers, specifically people who are educated both in the clinical and research domains. In this paper I use invention and patent data from two large Academic Medical Centers with over $1.2 billion in combined research revenue in 2011. My goal is to understand whether inventions created by inventors with boundary bridging, cross-domain[2] educational background are at a higher hazard of licensing. I use licensing as a proxy for translating an invention into a product.

I find that, contrary to expectations, inventions created by teams with cross-domain expertise– i.e. a combination of clinical and research education as proxied by an MD, PhD or MD/PhD degree in team members, have a significantly lower hazard of licensing compared to inventions by teams that are made up solely of clinicians or solely of bench researchers.

2.2 Introduction

Large amounts of money have been given in grants in recent years for translational research aimed at turning basic scientific knowledge into practical applications. Yet, we know

[1] See for example overview of Translational Research as part of the Clinical Research Roadmap Initiative at: http://commonfund.nih.gov/clinicalresearch/overview-translational.aspx, accessed February 19, 2012

[2] I use the term domain to refer to the clinic or the research bench as the use of discipline could be confused by discipline within science for example. Cross-domain then refers to teams that combine inventors with clinical and research backgrounds. This is also different from the term cross-functional as it relates not to the type of job role one is occupying within a company but to the knowledge background of the inventor measured by their education.

little about what makes the process of translation successful. A recent study of the life cycle of translational research for medical interventions notes that the average "translation lag" – i.e. the median time from earliest publication or patenting to time of first licensed clinical use was 24 years with an interquartile range of 14 to 44 years (Contopoulos-Ioannidis et al., 2008; cf. Morris, 2011).

The NIH has placed an emphasis on "nurtur[ing] a cadre of well-trained multi- and interdisciplinary investigators and research teams" and "synergiz[ing] multidisciplinary and interdisciplinary clinical and translational research and researchers to catalyze the application of new knowledge and techniques to clinical practice at the front lines of patient care."[3] It has been hypothesized that collaboration between the clinical and research domains is necessary for the translation process because very often clinicians are not only users of research but they possess unique expertise and knowledge by virtue of their interaction with patients (Demonaco et al., 2006; Baldessarini, 1985). Such knowledge can be crucial in informing further research and understanding ways in which basic research can be used in the clinic.

This emphasis is also based on the finding that the "burden of knowledge" has increased over time as knowledge in a field has become "deeper," thus forcing researchers to specialize more narrowly in their fields (Jones, 2009; Baumol et al. 2009). Such narrow specialization, especially in the life sciences and in times when knowledge creation has increased tremendously, may prevent one from seeing the interdependencies among different scientific and clinical findings. Thus, integration of knowledge becomes necessary for translation of basic science into the clinic (Kelley et. al. eds and IOM, 1994).

[3] "Focus on NINR and the NIH Roadmap," http://www.ninr.nih.gov/NR/rdonlyres/2C476ABF-E1F7-4BA3-B6F5-06EDF71225DE/0/RoadmapFocusFINAL113006.pdf accessed Feb 19th, 2012.

Interdisciplinary teams and individuals trained in both the clinical and research domains have been thought to be the ones who can accomplish the enormous task of translating science into the clinic and have had significant support.[4] Of the 129 medical schools in the country, 109 had an MD/PhD program as of 1990 and the NIH had invested $400 million in its own MST (MD/PhD) program from its inception to 1990 (Kelley et. al. eds and IOM, 1994). In addition, research has shown that first time NIH R01 grant applicants with an MD/PhD are significantly more likely to receive funding than those with an MD only or a PhD only. And among those that obtained such first funding, MD/PhDs were significantly more likely to get a second such grant compared to single degree researchers (Dickler at al., 2007).

In this paper, I explore whether teams with cross-domain knowledge are indeed faster to translate their inventions into the clinic compared to single-domain teams. For this task I use a dataset of 691 patents assigned to two Academic Medical Centers (AMCs). I am testing hypotheses that inventions by cross-domain teams (teams with both an MD and a PhD or an MD/PhD inventor) are at a greater hazard of being licensed, controlling for the type of invention, the scope of the patent, the innovativeness of the ideas and the experience of the lead inventor. Licensing is a way for new inventions to become products and is my proxy for translations of inventions into the clinic.

Surprisingly, I find that cross-domain teams are at a significantly lower hazard of licensing their inventions compared to single domain teams. As expected, I find that the experience of the inventor is positive and significant and leads to faster licensing as does the size of the team.

[4] By interdisciplinary here I mean teams that combine clinical and research backgrounds i.e. are comprised of members with MD degrees and PhD degrees.

2.3 Theory and Hypotheses Development

The influence of team diversity on productivity has been an important topic in the management literature. Theoretical models of team diversity have shown that a randomly chosen group of diverse problem solvers can outperform a group of high-ability problem solvers (Hong and Page, 2004). Empirically, teams that combine individuals with different information backgrounds based on their functions in the organization, also called cross-functional teams, have been studied extensively in the management literature for over half a century (e.g. Bunderson and Sutcliffe, 2002; Keller, 2001; Ancona and Caldwell, 1992; Galbraith, 1973; Kanter, 1988). In an innovative organization it has been argued that their importance stems from the need to "integrate expertise", "obtain and use distributed information" and provide for "speedy interdepartmental transfer" (Edmondson and Nembhard, 2009). Such a need is said to arise because of the different access to information and models of the world with which different departments operate inside the organization (Lawrence and Lorsch, 1969).

The research on whether cross-functional new product development (NPD) teams are in fact more innovative and faster to deliver new products has resulted in mixed findings. Some studies of cross-functional teams have found no effect (Cady and Valentine, 1999), but the majority have found a negative effect of functional diversity on performance (McDonough, 2000; Horowitz and Horowitz, 2007; also see reviews by Bettenhausen, 1991 and Williams and O'Reilly, 1998). Most of the studies that find a positive effect of team functional diversity on performance have been able to do so through mediator and moderator variables relating to group processes, task and group environment.

One of the key studies in this research stream looks at both direct and indirect effect of functional diversity on team performance for new product development teams (Ancona and

Caldwell, 1992). It finds that diversity has a direct negative effect on performance as evaluated by managers, particularly on dimensions of innovation. Team members' own evaluation of team performance is also negatively related to team diversity. The indirect effect of team diversity in this study, however, was positive as rated by managers. This effect was mediated by external communications. Groups that had higher levels of external communication were more highly rated and functionally diverse groups were more likely to communicate more externally.

Diverse teams face many challenges. Conflicts between different departments and functions within an organization may exhibit themselves in a cross-functional team. Task conflicts can further exacerbate the performance of a team (Pelled et al., 1999). In-group and out-group stereotyping can hinder communication and problem solving (Bunderson and Sutcliffe, 2002; Rico et al., 2007).

Diverse teams face informational challenges as well – they are not always able to communicate with the other team members to share the diverse information that they bring. Empirical studies have shown that group members often share only information that is common to all group members and are willing to discount their own knowledge and experiences in order to conform to group beliefs or hierarchy (Sherif, 1936; Asch, 1955; Nembhard et. al., 2006). Furthermore, it is not possible to encode and share all knowledge that an individual possesses, which limits the ability to use all available information in the group for problem solving (Polanyi, 1967; von Hippel, 1994).

The theoretical advantage of diverse knowledge in problem solving could be realizable if it were possible to integrate the diversity within one person on the team. Conflicts would be alleviated and information sharing would be less costly because the person who possesses diverse knowledge can serve as a "translator" thus mitigating team based challenges. The effects

of intrapersonal diversity – i.e. diversity within one person because of the person's educational or job background - has been studied previously. One study shows that in top management teams (TMTs) intrapersonal functional diversity on the team has a positive effect on information sharing and unit performance while interpersonal diversity had a negative effect (Bunderson and Sutcliffe, 2002). Similarly, the proportion of multi-knowledge individuals has an indirect positive effect, through information sharing, on product innovativeness and a direct positive effect on time efficiency of new product development teams as rated by managers (Park et al, 2009).

These results are based on innovation in companies, however, and on team performance as defined by speed or "innovativeness" and judged by managers who have appointed the teams rather than by more solid measures of innovative output. Managers may select teams to match the type of problem – i.e. select teams with intrapersonal diversity if such diversity is expected to help problem solving. Such selection is impossible to control for, except in experimental settings. One study shows that selection, or self-selection in this case, can lead to such positive findings. It finds that people with a more varied experience select to be entrepreneurs but conditional on being entrepreneurs, those with more varied experience earn less (Astebro and Thompson, 2011). As all non-experimental studies, my study doesn't entirely do away with selection but I use various controls and robustness checks to alleviate such concerns.

Furthermore, innovation in companies may be different from academic innovation. University inventions have tended to be on average much more important and general – i.e. basic, than industry inventions as indicated by patent citation measures. (Henderson et. al., 1997) The results could be different when bridging cutting edge science in very specialized and

different domains such as the clinic and the research bench. Yet, this question has not been explored in the university technology transfer literature.

University patenting, licensing and technology commercialization have been studied extensively, mostly with aggregated survey data gathered by the Association of University Technology Managers (AUTM) and with patent and licensing data from MIT (Shane, 2002; Decheneux et. al, 2008; Nerkar and Shane, 2007; Thursby et.al. 2004). Most of the research, however, has been based on the importance of patent characteristics and citation based measures of innovativeness. For example, Nerkar and Shane (2007) find that more pioneering inventions and patents with a wider scope are more likely to be commercialized, once licensed. Elfenbeim (2004), using Harvard University inventions, finds that the hazard of licensing increases in the inventor's prior publications as firms are more aware of such research.

Inventing team characteristics, however, have been largely ignored. Similarly, Academic Medical Centers (AMCs) have not figured extensively in the research on University innovation. This is surprising as AMCs tend to receive almost three times as much federal funding as regular universities and a higher amount of licensing income from their technologies (AUTM, 2006). In 2007, for example, of the universities that received over ten million in licensing income, all but two were Medical Schools (AUTM, 2007).

Academic Medical Centers represent the ideal setting in which to explore the interaction of the clinical and research domains and its influence on the translation of inventions into products. As institutions, AMCs combine both clinical practice and research labs and are therefore expected to spur interdisciplinary research and attract individuals with cross-domain expertise such as MD/PhDs. Such a research setting lets me explore the influence of team

characteristics along with invention characteristics to understand the translation of new technologies to the market and from there to the clinic.

The main question that I am trying to answer in this paper is whether inventions by cross-domain teams are at a greater hazard of being licensed than inventions by teams where all members come from the same domain. Furthermore, I explore whether having at least one person with a dual domain expertise on the team provides an added benefit to just having team members from different domains working together. Individuals with dual domain expertise could be able to integrate the different needs and knowledge of their domains to come up with inventions that can more easily be introduced into the clinic.

In an interview, an inventor with an MD/PhD from one of the AMCs in my dataset mentioned that he was better able to understand how a particular product will be used in the clinic because of his clinical experience. He gave examples of his own inventions from before he became a clinician which, he claimed, had suboptimal designs because he did not have a good understanding of how clinicians would use the product. After getting his MD he was able to adapt the design or the specific features of the cutting edge devices that he was developing to be more useable in the clinic. He argued that this leads to his current inventions becoming products faster than inventions that he created before he received his medical degree. Had he just worked with an MD instead of getting such a degree himself, he was afraid he may not be able to understand the different clinical need as well as he does now.

These anecdotal experiences lead me to develop the following hypotheses that can be tested using the dataset that I have:

Hypothesis 1: *Teams with cross-domain expertise come up with inventions that are at a greater hazard of licensing than teams with single-domain expertise.*

Hypothesis 2: *Teams with an MD/PhD on the team come up with inventions that are at a greater hazard of licensing than teams with single-domain expertise and teams with cross-domain expertise lacking an MD/PhD.*

2.4 Data

To test the above hypotheses I use invention data from a technology licensing office (TLO) that is in charge of the intellectual property of two academic medical centers (AMCs). Each observation in the data is a patent that has been applied for and granted between 1977 and 2008. That data is supplemented with patent data from the United States Patent and Trademark Office (USPTO) and licensing data from the same TLO as well as inventor data from various sources described below.

2.4.1 Research Setting

The employees of both AMCs are required to submit an invention disclosure to the TLO if they believe they have conceived of an idea that is novel and has a potential for commercialization. In the last decade the disclosure filing has been simplified through the availability of online forms. Once this disclosure is received at the TLO it is reviewed by a case manager who completes any missing information. Extensive research is then conducted by the TLO regarding the novelty of this invention, the quality and the potential for commercialization. A decision is made about whether to file an application for one or more patents based on this invention disclosure and a patent attorney from an external firm is retained to conduct a patentability opinion and proceed with patent filing. The case manager, inventor and the outside firm then coordinate to make any further decisions on the filing with the USPTO or patenting authorities in different countries.

The TLO also actively researches potential licensees for the invention. This involves finding out what other firms are doing research in this field and the general state of the art in the relevant technological area. Often, companies that may be interested are contacted. Inventors can and do initiate contacts to help commercialization as well since they jointly receive 25% of all income above expenses that is generated by the patent and another 25% goes to their lab. A short description of the invention is also available online for interested parties who can independently search for technologies from this TLO. Firms interested in the invention sign a confidentiality agreement to learn more about it before they decide to take an option or a license.

2.4.2 Invention Data

In the analysis for this paper my level of observation is a patent that has been granted in the United States to one of these two institutions. It becomes at-risk for licensing from the patent's priority or provisional filing date.[5] This is the closest date to the invention disclosure date and is different from the listed filing date in over 70% of the cases in my dataset. This discrepancy results from the fact that many patent applications result in multiple patents – divisions, continuations and continuation-in-parts of the parent application.[6] The expiration of a patent is determined by the priority date which is the earliest filing date in cases where a provisional patent has not been filed.[7]

[5] A priority date is the date on which the first patent from a particular invention was filed. Later that first patent may be abandoned and other divisions or continuations filed from it. Note that after June 8, 1995 inventors were allowed to file provisional patents which provide them with a year to decide whether a patent is going to be filed or not. The provisional patent date is not the same as the priority filing date from which the patent term is calculated.

[6] Talks with TLO officers indicated that I should use the filing date rather than the patent grant date as they start marketing an invention soon after the provisional patent is filed.

[7] A patent expires 20 years from the first filing date if that filing date is after June 8th, 1995. If the patent was filed before Jun 8, 1995, the term is *either* 20 years from the first filing date *or* 17 years from the grant date whichever is longer. A patent that has expired cannot be licensed as it is free for anyone to use.

Each patent publication on the USPTO web site also contains information on the inventors that are associated with it. This information can differ from the invention disclosure form because if an invention disclosure is split in different patents, the participation of team members on each patent can differ. For example, one inventor could have participated in inventing one of the claims in one patent but none of the claims in another. As such, data on the inventors listed on the patent publication is more accurate than TLO inventor data. In addition, while scientific papers may contain numerous authors who contributed little to the final product, a patent inventor, by law, must have contributed "to the conception of the invention" i.e. to at least one of the claims of the patent. If an inventor is not named on the patent or if someone who did not contribute an idea is named, the patent could be invalidated.[8]

The USPTO web site also contains information on the assignee of the patent. This is particularly important in the cases where inventions are the result of collaboration between members of different institutions, such as other companies or universities. I exclude patents that have as assignees or co-assignees institutions outside of the two AMCs that I am considering. When multiple institutions are assignees, they each have the right to license the patent without consulting with the co-assignees and licenses done by a different entity would not necessarily be observed by me. Often, co-assignees sign inter-institutional agreements to market and license the technology together but that can delay the time to licensing. An indicator variable for "more than one assignee" would not be useful to control for inventions handled by different or multiple TLOs as the outside institutions would be different. However, I include patents that were assigned together to the two AMCs that are in my dataset as they are managed by the same TLO and I observe all their licensing agreements.

[8] For more information see: http://www.uspto.gov/web/offices/pac/mpep/documents/2100_2137_01.htm, Accessed January 14, 2010.

There are patents in this dataset that were invented through collaborative efforts with companies or that resulted from sponsored research agreements (SRA) where a company gave money or equipment to support the research of a specific inventor or a department. These SRAs can be very specific – i.e. designing a prototype of an invention or very broad – i.e. supporting a lab's research without interference in the research direction. Patents resulting from SRAs are excluded from my dataset for a few reasons. First, they have a high rate of licensing (95%), almost invariably to the company that sponsored the research. Second, the company that sponsors the research will sometimes insist on patent filing for inventions that they are interested in licensing even if the TLO would not have filed a patent otherwise and I am not able to ascertain when that is the case. Third, while most SRAs give right of first refusal to the company, some give them automatic licenses at terms decided at the time of SRA signing and I cannot separate out inventions resulting from such SRAs since I don't have the text of every SRA. After I exclude all joint or sponsored patents I end up with 691 patents for my dataset.

On the USPTO website I am also able to observe abandonment dates for patents for which the maintenance fees have not been paid. This happens if the TLO decides after the patent is granted that it is unprofitable for them to maintain it by paying the maintenance fees. Maintenance fees in the US are due 3.5, 7.5 and 11.5 years after the patent has been granted.[9] As abandonment is equivalent to expiration, an abandoned patent is not available for licensing. I supplement the USPTO data which is not always updated online by data from IP Thompson's Delphion website.[10]

[9] Fees for 2007 depending on the size of the entity that owns the patent are as follows: at 3.5 years after first patent grant - $465 or $930 for small and large entity respectively, at 7.5 years $1190 and $2380, at 11.5 years $1965 and $3930. When the fees are not paid, the patent is abandoned at the 4th, 8th or 12th year respectively. If fees are paid with an additional fine sometime between year 3.5 and 4, the patent is not abandoned.
Source: http://www.uspto.gov/web/offices/com/sol/notices/71fr32285.pdf

[10] See at: www.delpion.com accessed January, 2010

Additional patent data, specifically patent cites, was obtained from the NBER patent dataset (Hall et. al, 2001). International Patent Classification (IPC) codes for each patent were obtained from the PatStat database by the EPO.[11] Patent cites are different from research article cites in that when a focal patent cites a previous patent it means that it is building on the previous invention. It further means that the material protected in the previous invention is limiting the scope of the current patent because the current patent cannot lay a claim on it. Furthermore, cites are added to the patent not only by the entity that files the patent but also by the patent examiner. A patent that cites multiple patents is believed to be less pioneering and narrower. Previous studies have found that a more pioneering university patent is more likely to get licensed (Nerkar and Shane, 2007; Dechenaux, 2008). A patent that is cited multiple times is considered important and valuable (Hall, 2005).

However, it is also well known that inventors tend to cite their own prior patents as they build on their inventions. This would mean that an invention with a higher number of prior self-citations or future self-cites will indicate a stronger research portfolio of the inventor and a stronger protection of intellectual property. Licensing agreements often include multiple patents and inventions in the same area and by the same inventor and firms will often license subsequent inventions by an inventor if they already have licensed one, especially if it enhances the protection to the first patent. This will speed up the time to licensing of inventions that are part of a stronger research portfolio which can be positively correlated with patent prior art. As a result, the predictions on the influence of prior art will depend on what portion of it is based on self-citations. It is also possible that the "portfolio" effect and the "pioneering" effect will cancel each other.

[11]See http://www.epo.org/searching/subscription/raw/product-14-24.html for more information. Accessed January 2012

I use the number of IPC codes on each patent as a measure of its scope. A patent with a larger scope will be more likely to be licensed because it would give a better protection and will be more valued by a potential licensor. Lerner (1994) shows that patent scope is related to a higher valuation by VCs in funding new firms. He finds that a one standard deviation increase in patent scope results in a 21% higher valuation for the firm. A larger patent scope has also been shown to increase the hazard of licensing an invention and commercializing it once it has been licensed in the university licensing setting (Elfenbeim, 2004; Nerkar, 2007).

To control for technology type, I use the main USPTO assigned patent classes and categorize the patents into six main groups – drugs, molecular biology, surgery, chemistry, optics and electricity and other with other including only 6% of patents. Appendix 2.1 provides a description of patent classes that are included in each category.

An alternative control for technology type uses all the IPC codes assigned to each patent at the subclass level. I select the 15 IPC codes at the subclass level that most frequently appear in my dataset. The 691 patents in my dataset have altogether 2074 IPC codes with 62 unique values at the subclass level. The top 16 of these unique values by frequency account for 91% of all IPC subclass level codes. I create a dummy variable for each of these 16 subclass level IPC codes. A 17^{th} category of "other" is a dummy for patents that do not belong to any of the 16 categories above – there are 33 patents in my dataset whose IPCs belong exclusively to the "other" category. Because a patent can have multiple IPC codes, a patent can belong to different categories as defined above.

This classification is important because the probability of licensing and the time to licensing depend on the type of technology being considered. For example, a medical device may be licensed faster than a drug molecule because it has a different clinical trial and FDA approval

timeline and return on investment. Furthermore, different types of firms will consider devices and drugs. Even within the large category of medical devices, some may be very different in terms of licensing than others – for example, software for an MRI machine will most certainly differ in licensing probability from a new catheter. Similarly, a research tool such as a genetically altered mouse may have an entirely different licensing profile than a drug molecule. Measuring these using various patent class measures provides further robustness. Since patented inventions are by definition unique, there is no classification that can account for all of the different patent characteristics. In classifying them, I am interested only in characteristics that will alter their licensing probability and timeline.

In addition to the patent classification using IPC classes, I have also classified them into types of claims by reading through the patent claims for each of the patents. I have three main categories -- process, object and combined process and object. It is possible that an object such as a new chemical composition will likely provide different strength of protection to a patent than a process claim for a method of treatment of a disease. Such level of protection then may influence whether and how soon these patents are licensed. These three categories are mutually exclusive.

I have further divided the patents into device and non-device inventions. A medical device is an apparatus that has human body contact and would require an FDA approval as a medical device. Even though there is a wide variation within medical devices with regard to their complexity and novelty, this classification is nevertheless useful as the approval process is quite similar for most devices and the types of firms that license these inventions are different from the firms that license non-device inventions.

2.4.2 Inventor Data

To test the importance of cross-domain experience on the team on licensing outcomes I use inventor educational background as an indicator of team diversity. It departs from previous literature where cross-functional teams have been classified as such based on different functional positions in the firm. Part of the reasoning has been that people with different functional positions would have different goals in addition to having different outlook on the world and these goals may be conflicting and may hinder team performance (Lawrence and Lorsch, 1969). In this case, I am interested in team composition as it relates to the different types of knowledge team members bring to the team. Because these inventors are all at top academic medical centers, they have one similar goal – to advance the knowledge of the scientific community and publish academic articles. I assume that many MDs will also have a goal to serve their patients and people with MD/PhDs will have a goal to enrich their clinical work with their research work and their lab work with the knowledge that they get from the clinic. Note however, that while MDs and MD/PhDs could choose to devote their career exclusively to research, PhDs do not have access to the clinic.

To find the educational background of inventors I use a number of sources described below. In their invention disclosures inventors are supposed to disclose their educational degree which is then on record with the TLO. More than half of the information about the inventors' background came from that file. However, because a few mistakes were discovered in that file, the data was further checked against other sources.

For every single inventor, a search was done through the AMCs internal directories. Unfortunately, the directory only has current information, so the rest of the degree information came from web sources. Each inventor's name was searched through the ProQuest Dissertations

and Theses Database, accessed through the Harvard University Library System.[12] A search was then conducted for every name using Google. CVs or university biography pages were used when available. Often, academic articles also include a researcher's degree and that information was very useful by letting us compare the date of the article and degree with the date of the invention. Sites such as vitals.com and healthgrades.com provide information on physician's medical education as well as year of graduation and licensing and were also a good, but last resort, source of information.[13]

Having information on graduation dates was useful in cases where a person has a patent and an article before getting an MD or PhD. In a few cases there were people who were already in an MD or a PhD program at the time of patent filing and they were considered MDs or PhDs respectively. Inventors that had only a Master's degree or a Bachelor's degree were classified as "other". For 11 of those, we were not able to ascertain the educational degree that the inventor had. In the dataset of 691 patents there are 1505 inventors because many patents have multiple inventors.

Team size is an important variable to control as it is potentially correlated both with team type and licensing outcome and can bias results. Team size is related to team type because teams with a higher number of inventors are more likely to be diverse. Inventions created by one inventor only, for example, cannot combine cross-domain expertise unless that cross-domain expertise is in one person. This relationship holds in my dataset as well – teams with a larger number of inventors are more likely to be cross-domain teams. Previous research has also found that patents with multiple inventors are cited more, implying that they are of a higher quality and as such may be more likely to get licensed (Wuchty et. al., 2007). One reason why licensing may

[12]See http://www.proquest.com/en-US/catalogs/databases/detail/pqdt.shtml, accessed 2011

[13]See http://www.vitals.com and http://www.healthgrades.com, accessed 2011

be affected by the number of inventors on the team may be that the more inventors there are on the team the more possible links there are to potential licensees through inventors' industry contacts. Tables 1 and 2 provide descriptions of team size and inventor educational degree

Table 2.1 Frequency distribution of number of inventors in team

Team size	Number of Patents
1	215
2	252
3	153
4	43
5	19
6	5
7	3
9	1
Total	691

Table 2.2 Number of Inventors by Educational Degree

Inventor Degree	Number of Inventors
MD/PhD	266
MD Only	570
PhD Only	578
Other Only	80
Degree Information Not Available	11
Total	1505

Because about 70% of patents have more than one inventor on them I need to devise a composite measure for team type based on the educational background of the inventors. To test my hypotheses, I divide my patents into five main groups – *Cross-Domain Integrated*, *Cross-Domain Distributed*, *Single-Domain Research*, *Single-Domain Clinical* and *Other*. *Cross-Domain Integrated* teams are those that have people who possess knowledge of both the clinical and research domain based on their educational background and that knowledge is integrated in at least one person of the team. This implies that the team has at least one inventor with an MD/PhD degree (two domains integrated in one person) but it may have additional MD/PhDs, or MDs or PhDs on the team. *Cross-Domain Distributed* teams have at least one MD and at least one PhD on the team but no inventor with an MD/PhD. All team members in *Single Domain Research* teams have PhDs (also can have a team member with an "other" degree in addition to PhDs) and all team members in *Single Domain Clinical* teams are MDs (also can include a member with an "other" degree in addition to MDs). All *Cross-Domain* and *Single-Domain*

teams can have researchers that are not PhDs or MDs. However, teams that have no MDs or PhDs are classified as "Other."

It is important to note that we don't have educational information for 11 inventors, four of these inventors belong to *Cross-Domain Integrated* teams and as such those inventions don't need to be dropped. Similarly, because *Cross-Domain* and *Single-Domain* teams can have members with *Other* degrees, we have 80 such inventors and only 6 teams that are in the *Other* category – i.e. have all of their members in the *Other* education category. Table 3 below contains a frequency distribution of the types of teams in the dataset.

Table 2.3 Frequency Distribution of Patents by Team Type

Team Type	Number of Patents
Cross-Domain Integrated	226
Cross-Domain Distributed	128
Single-Domain Clinical	173
Single-Domain Research	151
Other	6
At Least One Inventor with Unknown Degree Information and NOT Cross Domain Integrated	7
Total	691

For each patent, I also find out the lead inventor as noted by the TLO. The lead inventor (LI) is generally the Primary Investigator of the lab or the person with the most contributions to the patent. The lead inventor is also the one who participates most actively in patent prosecution and often has contacts with industry which may help during the licensing process. Some lead inventors have gone through the process of patent prosecution and licensing with previous inventions. While I don't observe patenting at other organizations by these inventors, this variable is still important as, at the very least, it measures the experience of the lead inventors working with this TLO. There are a few patents in my dataset for which the lead inventor is

missing or is not on the patent itself. This happens when an invention is disclosed and is then split into multiple patents and while one person has contributed the most to the whole invention, he is not an inventor on a specific patent. That person still remains the lead inventor for all inventions on the case. Patents like these are dropped in some of the analyses where lead inventor variables are included.

2.4.3 License Data

The information on licenses was extracted from the TLO's database. Patents may have more than one license associated with them – some are licensed through non-exclusive agreements, some are sublicensed by the first licensor. Additionally, many licenses are terminated and the invention then gets relicensed to a different entity. For my study I select the date of the earliest license for each patent in my dataset. Even if the first license gets terminated I consider licensing to be a proxy for attempt at commercialization or use in industry.

Of my 691 patents 358 (51%) are licensed at least once. Below is a table of the frequency distribution of licensing by team type.

Table 2.4 Licensing by Team Type

Team Type	Unlicensed Patents	Licensed Patents	Total
Cross-Domain Integrated	114	112	226
Cross-Domain Distributed	62	66	128
Single-Domain Clinical	71	102	173
Single-Domain Research	78	73	151
Other	5	1	6
Undefined	3	4	7
Total	333	358	691

2.5 Models and Results

The data for this study is right censored – I stop observing many patents that are still available for licensing at the end of the data-gathering process in early 2011. Furthermore,

because inventions are disclosed at different times, patents that were filed in the 1970s are observed until abandonment or expiration while later patents I observe for only some of their life. To account for all peculiarities of the data I use survival analysis models.

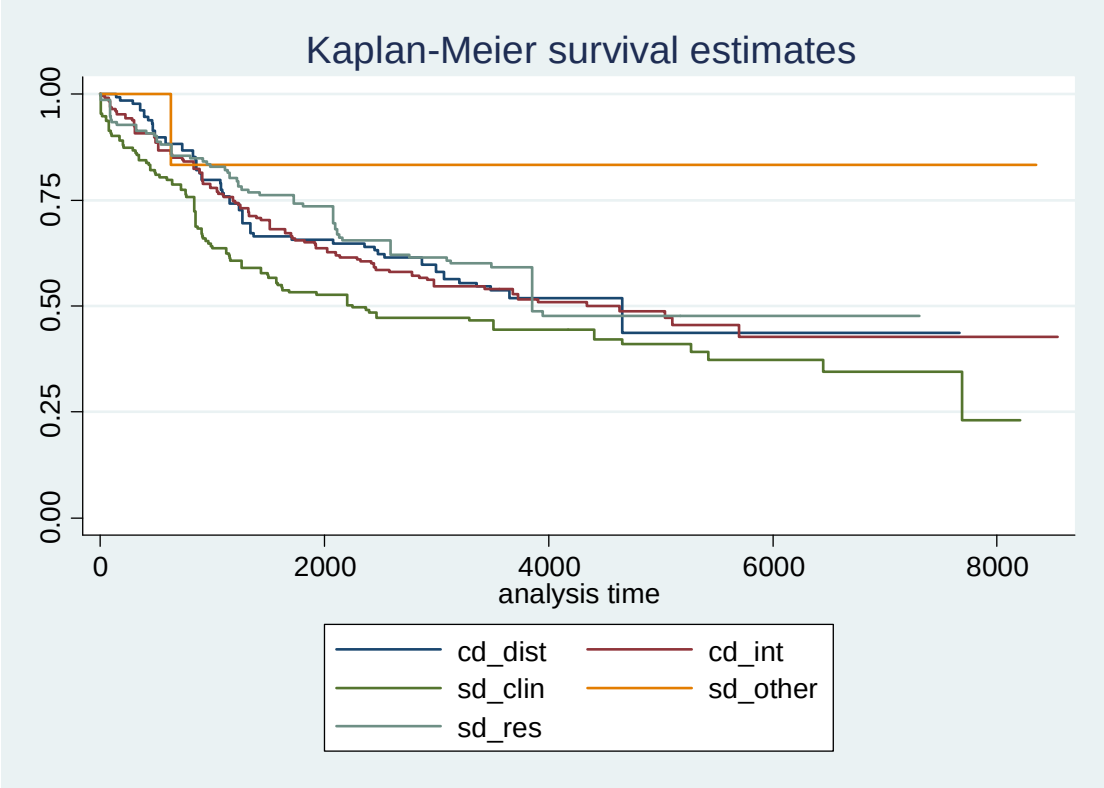
Survival analysis looks at time to licensing but accounts for right censoring as well as for inventions that are only in the data for a short period of time but do not achieve a license. I use parametric and non-parametric survival analysis models as well as random effects and shared frailty by lead inventor. I also do robustness checks by using standard models such as OLS of time to license and logit of probability of licensing within 10 years of first patent filing on case.

The simplest way to get some understanding of the distribution of survival data is to describe it graphically especially by separating it into the groups that we are interested in comparing. Below is a plot of the Kaplan-Maier estimate which is a non-parametric maximum likelihood estimate of the survivor function or the probability of survival past time *t*. It is given by the formula

$$\hat{S}(t)=\prod_{j|t_j\leq t}\left(\frac{n_j-d_j}{n_j}\right)$$

where n_j is the number of patents at risk at time t_j and d_j is the number of failures (licenses) at time t_j.

Graph 2.1 Kaplan – Meier Survival Estimates by Team Type



We note from the above graph that *Cross-Domain Integrated* (label: cd_int) and *Cross-Domain Distributed* (label: cd_dist) inventions follow similar survival curves while the *Single Domain* teams are different. Inventions by *Single Domain Clinical* (label: sd_clin) teams seem to get licensed faster than the rest, especially when they first become available. Inventions by *Single-Domain Research* teams, on the other hand, seem to be the slowest to license.

Simple statistics of the dataset are also reported based on the survival function estimate. The median, based on the survival function is the time after which only 50% of the sample survives (i.e. remains unlicensed) and the mean is the time that the average patent survives. Compare that to a raw calculation of the mean time of each patent in the sample without correction for censoring in the last column. Various test of equality of the survivor functions of

the different team types show that they are statistically significantly different. The result is robust to excluding the "Other" team type.

Table 2.6 Sample descriptive statics of survival by team type based on the Kaplan-Maier nonparametric estimate of the survival function. Time is reported as number of days

Team Type	Num of Observations	Median (50% of Sample Survives This Long)	Mean (The Average Patent Survives This Long)	Time in Sample (not calculated from survival function)
Cross Domain Integrated	226	4339	4747.456(*)	2903
Cross Domain Distributed	128	4657	4428.759(*)	3142
Single Domain Clinincal	173	2248	3856.621(*)	2638
Single Domain Research	151	3851	4469.332(*)	3082
Other	6	n/a	7063.167(*)	5703
Whole Sample	684	3851	4586.775(*)	2937

* largest observed analysis time is censored, mean is underestimated

Unfortunately, the non-parametric models shown above do not allow one to control for multiple variables. To be able to include the covariates into the regression, I use a semi-parametric proportional hazard model – the Cox Hazard Model. It estimates a baseline non-parametric hazard function – $h_0(t)$.

$$h(t|x_j) = h_0(t)e^{b_x x_j}$$

The covariates however enter the model linearly, i.e. they modify the hazard function multiplicatively, giving it the name proportional.

Table 2.7a-c include results from a Cox Proportional Hazard with no lead inventor controls (2.7a), models stratified by lead inventor (2.7b) and random effects models by lead inventor (2.7c).The random effects and stratification models are included to control for a potential omitted variable that is correlated both with the type of the team and the type of the

invention. It is possible, for example, that certain inventors select to work on a particular invention type – a clinician may work mostly on medical devices while most PhD teams work on molecular biology problems. Also the lead inventor may choose people to contribute to the team based on the team needs and certain lead inventors may be better at this selection than others. At the same time, we expect that different types of inventions will have a different hazard of licensing. It becomes particularly important then to control for invention type to avoid a selection bias.

It can be argued that team composition is not entirely endogenous in our sample. For one thing, the inventors on each particular patent are not determined by the lead inventor or by self-selection because of the nature of inventor determination at the USPTO. A team member can only be named an inventor on a patent if she contributed to the conception of the invention as mentioned above, not just reducing the idea into practice. As one invention turns into multiple patents, different team members may end up as inventors on different patents depending on how the various invention claims get divided between different patents. Including the above controls, however, is still important.

As discussed in the previous section, I use various invention type controls – I divide inventions into various categories based on their main USPTO classes. Those classes are described in Appendixes 2.1 and 2.2. Furthermore, I select the top 16 IPC codes and create categorical variables for whether a certain patent belongs to the particular IPC code. I further read and classify the patents into devices and non-devices based on whether they would require FDA approval for device. And lastly, I categorize patents into different groups based on whether they claim a process (method) or an object (composition of matter, article of manufacture or apparatus) or both.

While none of these are perfect invention type controls, they are different ways of addressing the issue of potential inventor selection into different types of technologies. An even better test would be to assume that each lead inventor creates a specific type of invention with its own different hazard of licensing. The Cox model allows us to control in two ways for lead inventor selection of invention type and team type - by using random effects by inventor and by stratifying by inventor.

Strata in a Cox Hazard model are disjointed groups based on certain characteristics – in this case, the lead inventor. Stratified models estimate a separate baseline hazard function for each stratum. They then assume that the covariates influence each stratum equally – the variable "device" influences the hazard function of lead inventor A and B in the same way. Shared frailty models (or the equivalent of random effects) on the other hand estimate the same baseline hazard function for all values of the lead inventor variable. They then assume that there is a latent variable that modifies the hazard function multiplicatively and has a different value for each group, in our case each lead inventor.

The stratified models are the preferred ones for this paper as they control for invention type in the way that I expect the data to behave. However, they exclude all patents of inventors who have only one invention. For that reason I report both the models stratified by lead inventor and the random effects models.

Table 2.7 presents the main results. In the first, third and fifth columns I compare the licensing hazard of all cross-domain integrated and cross-domain distributed teams to all single domain teams (the omitted category) using a regular Cox hazard model, a Cox Hazard model stratified by lead inventor and a random effects by lead inventor Cox hazard model respectively.

Table 2.7 Cox Hazard Model Results of licensing. USPTO Main Patent Class based technology controls. All standard errors clustered by invention disclosure and reported under coefficients. LI=Lead Inventor

			Stratified by LI	Stratified by LI	LI Random Effects	LI Random Effects
Cross-Domain Integrated	-0.409**	-0.024	-1.042**	-0.865*	-0.585**	-0.458*
	0.190	0.213	0.436	0.481	0.256	0.279
Cross-Domain Distributed	-0.497**		-0.263		-0.250	
	0.230		0.281		0.252	
Single-Domain All		0.338		0.075		0.022
		0.222		0.267		0.234
Patents Cited	0.0125***	0.0129***	0.0182***	0.0187***	0.0242***	0.0248***
	0.004	0.004	0.005	0.005	0.006	0.006
Forward Citations	0.636***	0.636***	0.529***	0.534***	1.064***	1.078***
	0.232	0.232	0.192	0.190	0.248	0.250
Patent Scope	0.0416*	0.0397*	0.0748*	0.0748**	0.0813**	0.0821**
	0.022	0.022	0.039	0.038	0.032	0.032
Device	-0.038	0.171	0.838***	1.085***	0.322	0.645**
	0.293	0.267	0.310	0.260	0.336	0.304
Device X Cross-Domain Integrated	1.121**	0.903**	0.746	0.445	0.788	0.445
	0.452	0.442	0.689	0.664	0.596	0.580
Device X Cross-Domain Distributed	0.954**		0.824*		1.336***	
	0.468		0.448		0.508	
Team Size	0.103	0.101	0.049	0.035	0.105	0.088
	0.066	0.066	0.110	0.109	0.076	0.076
Patent Filed Pre-1990	0.267	0.302	-0.932**	-0.856**	-0.231	-0.176
	0.253	0.252	0.401	0.408	0.288	0.288
Patent Filed 1990-1999	0.266	0.291	-0.387	-0.340	0.024	0.050
	0.224	0.224	0.335	0.339	0.245	0.245
Process Claims Only	0.184	0.178	0.591***	0.571***	0.413*	0.392*
	0.174	0.173	0.154	0.153	0.212	0.210
Process and Object Claims	-0.058	-0.056	0.167	0.138	-0.076	-0.087
	0.164	0.165	0.156	0.154	0.200	0.198
Molecular Biology Patent	0.740**	0.714**	1.286**	1.047**	1.218***	1.128**
	0.367	0.362	0.514	0.446	0.437	0.443
Drug Patent	0.671*	0.625*	1.422***	1.201***	0.935**	0.847**
	0.347	0.341	0.513	0.431	0.426	0.431
Chemistry Patent	0.865**	0.839**	1.826***	1.588***	1.613***	1.522***
	0.380	0.374	0.510	0.439	0.443	0.451
Surgery Patent	0.156	0.129	1.319**	1.104**	0.754*	0.656
	0.357	0.357	0.532	0.473	0.456	0.463
Optics/Electric Patent	-0.561	-0.645	0.749	0.556	-0.409	-0.576
	0.491	0.483	0.767	0.931	0.659	0.666
Observations	684	684	684	684	684	684
Num. of Inv. Disclosures					296	296

*** p<0.01, ** p<0.05, * p<0.1

Columns two, four and six compare cross-domain integrated teams to all other teams again using the respective models above. All standard errors are clustered by invention disclosure since multiple patents can result from one invention disclosure and the licensing probability of patents from the same invention disclosure may be correlated as firms often choose to license portfolios.

The models in which the omitted variable is all single-domain teams test our first hypothesis that all cross-domain teams' inventions are at a greater hazard of licensing than single-domain teams' inventions. The models where the omitted variable is cross-domain distributed teams test our second hypothesis that inventions by cross-domain integrated teams are at a higher hazard of licensing compared to inventions by cross-domain distributed teams.

The results are contrary to expectations –inventions by teams with cross-domain expertise are at a lower hazard of licensing than inventions by single-domain teams contradicting our first hypothesis. This result is, in fact, driven by cross-domain integrated teams whose inventions are also at a lower hazard of licensing when compared only to inventions by cross-domain distributed teams, which also contradicts our second hypothesis. Furthermore, the latter effect is significant in all models except for the models with no lead inventor group controls.

It is important however, that when we interact the cross-domain integrated and cross-domain distributed variables with the dummy variable "device" we always get positive and at times statistically significant results. In fact, the magnitude of that coefficient is sometimes larger (also in absolute value) than the magnitude of the coefficient on the team type variables. This indicates that while on average inventions by cross-domain teams are at a lower hazard of licensing, when those inventions are medical devices, they may in fact be at a higher hazard of licensing than non-device inventions created by single-domain teams.

Table 2.8 Cox Hazard Model Results of licensing. IPC Subclass based technology type controls. All standard errors clustered by invention disclosure and reported under coefficients.

			Stratified by Lead Inventor	Stratified by Lead Inventor	Lead Inventor Random Effects	Lead Inventor Random Effects
Cross-Domain Integrated	-0.366*	0.020	-1.134**	-0.859*	-0.512**	-0.331
	0.201	0.222	0.456	0.513	0.255	0.282
Cross-Domain Distributed	-0.503**		-0.330		-0.311	
	0.236		0.324		0.259	
Single-Domain All		0.335		0.220		0.087
		0.225		0.309		0.242
Patents Cited	0.00684**	0.00695**	0.020***	0.021***	0.022***	0.023***
	0.003	0.004	0.006	0.006	0.006	0.007
Forward Citations	0.639***	0.652***	0.429**	0.439**	0.935***	0.962***
	0.217	0.218	0.188	0.185	0.254	0.255
Patent Scope	0.102	0.092	-0.100	-0.103	0.021	0.014
	0.119	0.119	0.131	0.134	0.109	0.110
Device	-0.472*	-0.240	0.516*	0.680***	-0.210	0.129
	0.275	0.261	0.305	0.256	0.336	0.305
Device X Cross-Domain Integrated	1.135**	0.896**	0.537	0.374	0.770	0.431
	0.467	0.454	0.646	0.638	0.572	0.556
Device X Cross-Domain Distributed	1.004**		0.496		1.286***	
	0.464		0.527		0.497	
Team Size	0.096	0.091	0.069	0.057	0.096	0.082
	0.069	0.068	0.119	0.118	0.078	0.078
Patent Filed Pre-1990	0.243	0.269	-0.901*	-0.859*	-0.139	-0.098
	0.269	0.269	0.482	0.500	0.288	0.288
Patent Filed 1990-1999	0.288	0.305	-0.367	-0.338	0.050	0.067
	0.225	0.227	0.379	0.381	0.243	0.242
Process Claims Only	0.127	0.125	0.266*	0.263*	0.108	0.099
	0.173	0.174	0.139	0.139	0.207	0.206
Process and Object Claims	-0.046	-0.037	-0.056	-0.069	-0.295	-0.301
	0.167	0.167	0.141	0.140	0.186	0.185
IPC Subclass Dummies (Top 16 by Frequency)	Included	Included	Included	Included	Included	Included
	Some Significant	Some Significant	Some Significant	Some Significant	Some Significant	Some Significant
Observations	684	684	684	684	684	684
Number of groups					296	296

*** p<0.01, ** p<0.05, * p<0.1

The negativity of the results is robust to multiple specifications of invention type based on patent measures – i.e. USPTO classes vs. IPC subclasses. The only exception is the comparison of cross-domain integrated to cross-domain distributed teams with the IPC based measure in the model without controls for lead inventor groups. The coefficient on our cross-domain team variable there is positive but very small and highly insignificant. Recall that patents can belong to multiple IPC subclasses but only one USPTO class. Please see Table 2.8 above for comparisons.

Tables 2.9 and 2.10 go on to explore the results further by including only patents that have more than one inventor. Team size is correlated with team type because patents that have only one inventor cannot belong to the cross-domain distributed teams. Furthermore, the use of cross-domain expertise may be different in single inventor teams and teams with multiple individuals who each have a different knowledge background.

The results in Table 2.9 are largely consistent with our findings from the whole sample models in Table 2.7 but statistical significance is weaker. For example, in the models without lead inventor controls, when cross-domain integrated teams are compared to cross-domain distributed teams (column 2), the coefficient is now positive even with USPTO class controls, albeit very small and statistically insignificant. Similarly, in the first model, where both distributed and integrated cross-domain team inventions are compared to the single-domain team inventions, the coefficient on the cross-domain integrated team variable is not statistically significant any more at any of the generally accepted levels.

Table 2.9 Cox Hazard Model Results of licensing. USPTO Main Patent Class based technology controls. **Sample excludes sole inventor patents.** All standard errors clustered by invention under coefficients. LI=Lead Inventor

			Stratified by LI	Stratified by LI	LI Random Effects	LI Random Effects
Cross-Domain Integrated	-0.368	0.022	-1.197**	-1.202**	-0.720**	-0.573*
	0.234	0.215	0.503	0.541	0.311	0.316
Cross-Domain Distributed	-0.488*		-0.077		-0.277	
	0.259		0.309		0.302	
Single-Domain All		0.301		-0.164		-0.025
		0.248		0.299		0.274
Patents Cited	0.0109***	0.0113***	0.0190***	0.0203***	0.0272***	0.0286***
	0.004	0.004	0.007	0.007	0.007	0.007
Forward Citations	0.448*	0.438*	0.348*	0.321	0.852***	0.843***
	0.265	0.264	0.202	0.199	0.305	0.307
Patent Scope	0.0488*	0.044	0.0659*	0.0644*	0.057	0.055
	0.028	0.028	0.040	0.038	0.038	0.038
Device	-0.056	0.241	0.669*	1.044***	0.236	0.703*
	0.373	0.323	0.370	0.317	0.420	0.364
Device X Cross-Domain Integrated	1.052*	0.737	1.700**	1.208	1.395**	0.881
	0.557	0.528	0.766	0.748	0.708	0.672
Device X Cross-Domain Distributed	0.920*		0.864		1.371**	
	0.518		0.542		0.573	
Team Size	0.058	0.067	0.087	0.091	0.134	0.126
	0.087	0.084	0.127	0.128	0.098	0.099
Patent Filed Pre-1990	0.219	0.260	-0.806*	-0.715	-0.113	-0.059
	0.280	0.279	0.438	0.438	0.336	0.337
Patent Filed 1990-1999	0.112	0.120	-0.098	-0.069	-0.001	0.018
	0.241	0.242	0.331	0.336	0.279	0.278
Process Claims Only	0.160	0.154	0.578***	0.586***	0.371	0.377
	0.219	0.218	0.194	0.196	0.258	0.256
Process and Object Claims	-0.056	-0.052	0.158	0.150	-0.017	-0.002
	0.212	0.213	0.174	0.178	0.240	0.238
Molecular Biology Patent	0.782*	0.748*	1.759*	1.451*	1.481***	1.346**
	0.460	0.453	0.907	0.869	0.568	0.573
Drug Patent	0.775*	0.709	1.795*	1.494*	1.144**	0.984*
	0.441	0.432	0.928	0.877	0.561	0.564
Chemistry Patent	1.114**	1.062**	2.318***	2.009**	2.220***	2.039***
	0.482	0.471	0.879	0.835	0.577	0.584
Surgery Patent	0.437	0.413	1.463*	1.206	1.263**	1.116*
	0.442	0.446	0.857	0.840	0.593	0.598
Optics/Electric Patent	-0.133	-0.221	1.034	0.788	0.220	-0.033
	0.626	0.624	0.999	1.091	0.814	0.817
Observations	469	469	469	469	469	469
Number of groups					227	227

*** p<0.01, ** p<0.05, * p<0.1

Table 2.10 Cox Hazard Model Results of licensing. IPC Subclass based technology type controls. **Sample excludes sole inventor patents.** All standard errors clustered by invention disclosure and reported under coefficients.

			Stratified by Lead Inventor	Stratified by Lead Inventor	Lead Inventor Random Effects	Lead Inventor Random Effects
Cross-Domain Integrated	-0.292	0.126	-1.140**	-1.067**	-0.467	-0.285
	0.238	0.223	0.498	0.517	0.294	0.305
Cross-Domain Distributed	-0.510*		-0.124		-0.270	
	0.263		0.361		0.293	
Single-Domain All		0.331		0.005		0.072
		0.249		0.354		0.274
Patents Cited	0.006	0.006	0.0283***	0.0289***	0.0240***	0.0256***
	0.004	0.004	0.009	0.009	0.008	0.008
Forward Citations	0.484*	0.500**	0.240	0.229	0.739**	0.749**
	0.248	0.247	0.187	0.185	0.305	0.306
Patent Scope	0.197	0.198	-0.068	-0.060	0.058	0.072
	0.191	0.195	0.147	0.154	0.155	0.157
Device	-0.485	-0.142	0.330	0.580**	-0.167	0.243
	0.370	0.331	0.353	0.264	0.421	0.356
Device X Cross-Domain Integrated	1.225**	0.877*	1.278*	1.042	1.377**	0.978
	0.523	0.497	0.764	0.733	0.661	0.627
Device X Cross-Domain Distributed	0.910*		0.493		1.020*	
	0.498		0.538		0.543	
Team Size	0.073	0.075	0.136	0.138	0.100	0.100
	0.090	0.088	0.147	0.148	0.098	0.099
Patent Filed Pre-1990	0.127	0.161	-0.821	-0.753	-0.094	-0.055
	0.306	0.307	0.533	0.552	0.332	0.333
Patent Filed 1990-1999	0.142	0.144	-0.218	-0.190	-0.030	-0.024
	0.237	0.240	0.348	0.347	0.271	0.270
Process Claims Only	0.110	0.118	0.282*	0.290*	0.138	0.156
	0.227	0.226	0.157	0.157	0.266	0.264
Process and Object Claims	0.005	0.011	0.035	0.021	-0.195	-0.194
	0.215	0.216	0.175	0.173	0.232	0.229
IPC Subclass Dummies (Top 16 by Frequency)	Included Some Significant	Included Some Significant	Included Some Significant	Included Some Significant	Included Some Significant	Included Some Significant
Observations	469	469	469	469	469	469
Number of groups					227	227

*** p<0.01, ** p<0.05, * p<0.1

Statistical significance of the team type variable coefficients disappears in most of the models in which the sample is restricted to teams with two or more inventors and in which IPC based controls are used. The coefficients on cross-domain team variables are not at all significant in the random effects models but still remain significant in the stratified models. The results imply that our negative results were driven largely by patents invented by cross-domain integrated teams with single inventors – i.e. lone inventors with MD/PhD degrees.

It is important to mention that the lack of significant negative results still does not support our hypothesis that inventions by cross-domain teams are at a higher hazard of licensing than single-domain team inventions or that inventions by cross-domain integrated teams are at a higher hazard of licensing than inventions by cross-domain distributed teams. It alleviates however the problem of explaining negative results that are counter to our stated hypotheses.

Of the control variables, one is particularly important to note – the number of patents cited by the focal patent which is an indicator of how pioneering the invention is. Inventions with fewer cited patents are more pioneering and according to previous study they are also more likely to be licensed (Nerkar et al., 2007). In our dataset that result does not hold. Inventions that cite more patents are at a higher hazard of licensing. As noted above, this could be due to the fact that inventions with more prior art may be citing patents by the same inventor which may imply a larger and potentially better protected patent portfolio. It could also imply, however, that industry needs time to catch up with university technology and more cutting edge patents take longer to be recognized. Further research to understand the mechanism behind this result is needed because of its implications for technology transfer policy.

2.6 Conclusion

In this paper, I try to understand the role of cross-domain expertise on patent licensing from Academic Medical Centers. My results are contrary to expectations implying that inventions by *Cross Domain* teams are at a lower hazard of licensing than those by *Single Domain* teams. These results are influenced by the type of technology implying that the importance of cross-domain expertise on the inventing team varies by various types of inventions and that identifying those areas and directing MD/PhDs into them may present efficiency gains. Potential educational subsidies for MD/PhDs specializing in degrees that would be associated with future device inventions, for example, could be considered. However, more research needs to be done in this area before such recommendations can be made.

Even if the results are predictive of trends in the larger population they do not imply causality as my teams and inventions are simultaneously determined – i.e. teams of different educational background are not randomly assigned to come up with a specific invention, instead people choose to work on a certain type of invention. Random assignment of inventors to teams that can then come up with a specific invention is not possible. Choice of area in which to work is personal and hard to control for. While I control for the technological class and subclass of a patent, there may be other characteristics of the invention that are correlated both with the inventor degrees and with licensing.

Furthermore, lower rates of licensing may not mean that MD/PhDs or cross-domain teams are not crucial to the translational process. Patents and patent licensing measure only a certain type of creative and translational activity. It is possible that cross-domain teams are better at translating existing inventions into the clinic rather than coming up with new licensable ones.

It could also be that even though their inventions are at a lower hazard of licensing, they may be faster to commercialize once licensed.

More research is needed to understand the role of cross-domain work, teams and individuals for the process of translating basic science into the clinic. This paper is providing a starting point and puzzling findings that can spur further investigation.

2.7 Bibliography

Ancona, D and D. Caldwell (1992), "Demography and Design: Predictors of New Product Team Performance." *Organization Science*, vol.3, pp. 321-341

Asch, S. E. (1951), "Effects of group pressure upon the modification and distortion of judgment," in H. Guetzkow (ed.), *Groups, leadership and men*, Pittsburgh, PA: Carnegie Press

Asch, S.E. (1955), "Opinions and Social Pressure", Scientific American, 193, pp. 31-35

Astebro T.B. and P. Thompson (2011), "Entrepreneurs: Jacks of all Trades or Hobos?" *Research Policy*, June 2011, vol. 40, no. 5, pp. 637-649

Association of University Technology Managers, (2006) "AUTM STATT Database," http://www.autm.net/source/STATT/index.cfm?section=STATT, accessed January 2009

Association of University Technology Managers, (2007) "AUTM STATT Database http://www.autm.net/source/STATT/index.cfm?section=STATT," accessed January 2009

Baldessarini R.J. (1985), Chemotherapy in Psychiatry: Principles and Practice (second edition), Cambridge, MA: Harvard University Press

Baumol,W.J., Schilling, M.A. and E.N. Wolff (2009), "The Superstar Inventors and Entrepreneurs: How Were They Educated?" *Journal of Economics & Management Strategy,* vol.18, no. 3, pp. 711- 728

Bettenhausen, K.L. (1991), "Five Years of Groups Research: What We Have Learned and What Needs to Be Addressed," *Journal of Management* vol.17, no.2, pp. 345–381

Bunderson, J.S. and K.M. Suthcliffe (2002), "Comparing alternative conceptualizations of functional diversity in management teams: Process and performance effects," *Academy of Management Journal*, vol. 45, no.5, pp. 875-893

Cady, SH and J. Valentine (1999), "Team innovation and perceptions of consideration – What difference does diversity make?," *Small Group Research*, vol. 30, no. 6, pp. 730-750

Chatterji, Aaron (2009) "Spawned with a Silver Spoon? Entrepreneurial Performance and Innovation in the Medical Device Industry" Strategic Management Journal vol. 30, pp.185-206

Contopoulos-Ioannidis, D.G., Alexiou, G.A., Gouvias, T.C. and J.P.A. Ioannidis JPA, Medicine (2008) – Life-cycle of translational research for medical interventions. *Science*, vol. 321, pp. 1298-1299

Dechenaux E., Goldfarb, B., Shane S. and M. Thursby (2008), "Appropriability and commercialization: Evidence from MIT inventions," *Management Science*, vol.54, no. 5, pp. iv-iv

Demonaco, H., Ali, A. H. and E. von Hippel (2006), "The Major Role of Clinicians in the Discovery of Off-Label Drug Therapies," *Pharmacotherapy,* vol. 26, no. 3, pp 323-332

Dennis, A. R. (1996), "Information exchange and use in group decision making: you can lead a group to information, but you can't make it think," *MIS Quarterly*, vol. 20, no.4, pp. 433-457

Dickler H.B., Fang, D., Heinig, S.J., Johnson, E. and D. Korn (2007) "New Physician-Investigators Receiving National Institutes of Health Research Project Grants: A Historical Perspective on the "Endangered Species," JAMA, vol. 297, no. 22, pp. 2496-2501

Edmondson, A.C. and I.M. Nembhard (2009), "Product Development and Learning in Project Teams: The Challenges Are the Benefits," *Journal of Product Innovation Management*, vol. 26, no. 2, pp. 123-138

Elfenbein, DW (2007), "Publications, patents, and the market for university inventions," *Journal of Economic Behavior & Organization, vol.* 63, no.4, pp. 688-715

Galbraith, J (1973), Designing Complex Organizations, Addison-Wesley: Reading, MA

Hall, B. H., Jaffe, A.B. and M. Trajtenberg (2001), "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," *NBER Working Paper 8498*

Hall, B., A Jaffe, and M Trajtenberg, (2005) "Market Value and Patent Citations," *RAND Journal of Economics, vol.* 36, no. 1, pp. 16-38

Henderson, R. M., Jaffe, A., and M. Trajtenberg (1997), "University versus Corporate Patents: A Window on the Basicness of Invention," *Economics of Innovation and New Technology, v.* 5, no. 1, pp. 19-50

Hong, L. and S.E. Page (2004), "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers," Proceedings of the National Academy of Sciences, pp.16385-16389

Horwitz S. K. and I. B. Horwitz (2007), "The effects of team diversity on team outcomes: a meta-analytic review of team demography, *Journal of Management,* vol. 33, no.3, pp. 987-1015

Jones, B. F. (2009) "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" *Review Of Economic Studies* 76, no. 1, pp. 283-317

Kanter, R. M. (1988), "When a Thousand Flowers Bloom: Structural, Collective and Social Conditions for Innovation in Organizations", In: Research in Organizational Behavior, ed. B. M. Staw, and L. L. Cummings. Greenwich, CT: JAI Press, 97–102

Keller, R.T. (2001), "Cross-Functional Project Groups in Research and New Product Development: Diversity, Communications, Job Stress, and Outcomes," *Academy of Management Journal*, vol. 44, no.3, pp. 547–555

Kelley, W. and Institute of Medicine (1994) Careers in clinical research: obstacles and opportunities, Washington D.C.: National Academy Press

Lawrence. O. and J. Lorsch (1969), "Differentiation and Integration in Complex Organizations," *Administrative Science Quarterly,* vol. 12, no.1, pp.153-167

Lerner J., (1994), "The Importance of Patent Scope - An Empirical-Analysis," *RAND Journal Of Economics,* vol. 25, no. 2, pp. 319-333

Lettl, C., Herstatt, C. , Gemuenden, H.G. (2006): Users' contributions to radical innovation: Evidence from four cases in the field of medical equipment technology. *R&D Management*, Vol. 36, 251-272

McDonough III, E.F. (2000), "Investigation of Factors Contributing to the Success of Cross-Functional Teams," *Journal of Product Innovation Management*, vol. 17, No. 3, pp. 221–235

Morris, Z.S., Wooding, S. and J. Grant (2011), "The answer is 17 years, what is the question: understanding time lags in translational research," Journal of the Royal Society of Medicine, vol. 104, no. 12, pp. 510-520

Nembhard, I.M. and A.C Edmondson (2006), "Making It Safe: The Effects of Leader Inclusiveness and Professional Status on Psychological Safety and Improvement Efforts in Health Care Teams," *Journal of Organizational Behavior*, vol. 27, no. 7, pp. 941–966

Nerkar, A. and S. Shane (2007), "Determinants of invention commercialization: An empirical examination of academically sourced inventions," *Strategic Management Journal,* vol. 28, no. 11, pp.1155-1166

Park, M.H.J., Lim, J.W. and P.H. Birnbaum-More (2009), "The Effect of Multiknowledge Individuals on Performance in Cross-Functional New Product Development Teams," *Journal of Product Innovation Management,* vol. 26, no.1 pp. 86-96

Pelled, L.H., Eisenhardt, K.M., and K.R. Xin. (1999), "Exploring the Black Box: An Analysis of Work Group Diversity, Conflict, and Performance," *Administrative Science Quarterly*, vol. 44, no. 1, pp. 1–28

Polanyi, M. (1967), The Tacit Dimension, Garden City, NY: Doubleday Anchor Books

Randel, A.E. and K.S. Jaussi (2003), "Functional background identity, diversity, and individual performance in cross-functional teams," *Academy of Management Journal* vol. 46, no.6, pp. 763-774

Rico, M., Molleman, E., Sanchez-Manzanares, M. and G.S. Van der Vegt (2007), "The effects of diversity faultlines and team task autonomy on decision quality and social integration," *Journal of Management*, vol. 33, no. 1, pp. 111-132

Shane, S. (2002), “Selling university technology: Patterns from MIT,” *Management Science,* vol. 48, no. 1, pp. 122-137

Sherif, M. (1936), *The psychology of social norms*, New York, NY: Harper

Thursby, JG and MC Thursby, (2004) “Are faculty critical? Their role in university-industry licensing,” *Contemporary Economic Policy,* vol. 22, no. 2, pp.162-178

Trajtenberg, M. (1990), “A Penny for Your Quotes - Patent Citations and the Value of Innovations,” *RAND Journal of Economics,* vol.21, No. 1, pp. 172-187.

Von Hippel (1994), "Sticky Information" and the Locus of Problem Solving: Implications for Innovation,” *Management Science*, vol. 40, no. 4, pp. 429-439

Williams, K.Y. and C.A. O’Reilly, III. (1998), “Demography and Diversity in Organizations: A Review of the 40 Years of Research,” *Research in Organizational Behavior* vol. 20, no.20, pp. 77–140

Wuchty S, BF. Jones, and B. Uzzi, (2007) “The Increasing Dominance of Teams in Production of Knowledge,” *Science*, vol. 316, no. 5827, pp.1036-1039

2.8 Appendixes

Appendix 2.1 USPTO Main Patent Classes, Classification into Larger Class Groups and Frequency of Occurrence in Data

Class Group	Main USPTO Class	Frequency in Dataset		Class Group	Main USPTO Class	Frequency in Dataset
Chemistry	506	2		Other	101	1
Chemistry	530	36		Other	198	1
Chemistry	536	25		Other	345	3
Chemistry	544	1		Other	564	1
Chemistry	552	2		Other	706	1
Chemistry	554	7		Other	5	5
Chemistry	436	12		Other	106	1
Chemistry	560	1		Other	210	2
Total Chemistry		**86**		Other	250	2
Drug	424	96		Other	340	1
Drug	514	145		Other	372	1
Total Drug		**241**		Other	422	2
Molecular Biology	435	125		Other	427	1
Molecular Biology	800	13		Other	433	2
Total Molecular Biology		**138**		Other	521	1
Optics/Imaging	356	6		Other	522	2
Optics/Imaging	359	2		Other	523	2
Optics/Imaging	378	5		Other	525	2
Optics/Imaging	382	3		Other	528	1
Optics/Imaging	385	3		Other	623	8
Optics/Imaging	702	3		Other	705	1
Optics/Imaging	324	20		**Total Other**		**41**
Total Optics/Imaging		**42**				
Surgery	128	14				
Surgery	600	56				
Surgery	601	7				
Surgery	604	18				
Surgery	606	37				
Surgery	607	11				
Total Surgery		**143**				

Appendix 2.2 IPC Codes Frequency of Occurrence in Sample Patents. A patent may have multiple IPC Codes

IPC Code at Subclass Level	Number of Occurrences		IPC Code at Subclass Level	Number of Occurrences
A61 K	657		C7 F	5
A61 B	250		C8 L	5
C7 K	248		C8 G	5
G1 N	135		G9 G	3
C12 N	123		A47 D	3
C7 C	115		B6 B	3
A61 F	58		B41 K	2
A1 N	50		C7 J	2
A61 M	49		B29 B	2
A61 N	46		A1 H	2
G1 R	44		C9 H	2
C12 Q	39		A47 G	2
A61 L	34		H4 N	2
G2 B	22		G21 K	1
C7 H	20		H4 Q	1
C12 M	17		C2 F	1
C7 D	13		A47 C	1
A61 G	9		G10 K	1
C12 P	9		B1 L	1
G6 T	9		B65 G	1
H4 L	8		G6 K	1
G6 F	7		G1 S	1
C8 B	7		B5 D	1
C7 B	7		G1 J	1
G1 T	6		B28 B	1
C40 B	6		G1 F	1
A61 Q	6		G1 V	1
A61 C	6		C9 D	1
A1 K	6		F2 B	1
B29 C	6		H1 F	1
G1 B	6		C7 G	1

3. DO DIVERSITY AND FOCUS IN ROUTINE WORK INFLUENCE CREATIVE OUTPUT? EVIDENCE FROM CARDIAC SURGERY

Ayfer Ali **Robert Huckman**

3.1 Abstract

Routine, non R&D work has rarely been seen as the wellspring of innovative ideas. In this paper we propose that the design of routine work can influence the quantity and quality of innovation. We use the setting of cardiac surgery to show that a surgeon's clinical focus – i.e. the quantity and type of procedures he performs -- has an influence on the quantity and quality of his academic article publications, which are our proxy for innovative output. We use a panel dataset of 162 surgeons who perform procedures at academic hospitals in New York State from 1994 to 2004. We find that performing a more diverse set of cardiac procedures is associated with a higher number of articles for early-career surgeons. We also find that as the percentage of heart valve procedures rises as part of the surgeon's portfolio, so does the number of cite-weighted articles related to heart valve procedures. This result does not hold for Coronary Artery Bypass Graft (CABG) procedures however – an increase in the percentage of CABG procedures that a surgeon performs, does not increase the number of cite-weighted CABG articles.

3.2 Introduction

Many workers have dual roles in their jobs. They must not only execute routine tasks efficiently to ensure short-term success, but they must also creatively alter and improve the way they perform those tasks over the long term. At the organizational level March (1991) refers to these dual roles as exploitation and exploration, respectively, and describes them as competing for organizational resources – an organization that spends time exploiting current expertise will

not spend that time exploring new ways of doing things. The view that exploitation and exploration represent fundamentally different activities, suggests that individuals or organizations that are geared to succeed at exploitation should be less proficient at exploration and vice-versa. At a minimum, one might expect that performance with respect to these two types of activities is uncorrelated.

In this paper, we consider whether the relationship between exploitation and exploration is more nuanced than simply being uncorrelated or being a "pure tradeoff". Specifically, we examine whether the manner in which an individual's routine work (i.e. exploitation) is structured influences his or her innovative performance (i.e. exploration). We characterize the routine work of an individual by its degree of specialization under the assumption that greater specialization i.e. repeatedly performing one type of task at the expense of others, might impede innovative performance to the extent that it does not expose the individual to significant variety (Levitt and March, 1988).

We examine this issue within the empirical context of cardiac surgery. We use this context because a relatively large percentage of cardiac surgeons are affiliated with academic medical centers, implying that they are responsible for both performing surgical procedures (akin to exploitation) and performing academic research (akin to exploration). With respect to the latter activity, cardiac surgeons at academic hospitals often innovate in terms of devising new techniques or medical devices (Riskin et al., 2006; Chatterji et al., 2008). Furthermore, they are part of the larger academic community that publishes articles on new discoveries and advances in their field and their innovative performance is reflected in their publications (Merton, 1957; Stephan, 1996). Ultimately, this context allows us to examine how the structure of an individual's exploitation activities impacts his or her performance in terms of exploration.

We find that the structure of routine work is indeed related to the innovative output of cardiac surgeons but only statistically significantly for the sample of early career surgeons. A more diverse set of procedures results in more articles published for early-career cardiac surgeons. Furthermore, innovation in certain areas (i.e. heart valve procedures) is associated with a higher focus in that area, while innovation in other areas (i.e. CABG procedures) is associated with a higher focus in related (cardiac but non-CABG) areas.

The remainder of this paper proceeds as follows. Section 3.3 develops our hypothesis. Section 3.4 describes the data and Section 3.5 shows our results. Section 3.6 concludes with a discussion of generalizability and limitations.

3.3 Literature Review and Hypothesis Development

The idea that knowledge can be gained through work is not new. Learning-by-doing has been shown to improve productivity (Wright, 1936; Arrow, 1962; Argote and Epple, 1990). Workers can improve performance in several ways such as by learning more-efficient ways to perform specific tasks, learning to use tools, or learning to work together more effectively (Edmondson et. al, 2003; Huckman and Pisano, 2006; Lapre and Nembhard, 2010). New problems that were not anticipated can become apparent in the course of work and may need to be resolved, resulting in modifications in processes and tools that improve productivity (Adler and Clark, 1991). Though individual modifications may be incremental, they can accumulate to become significant innovations with a large economic value (Rosenberg, 1979; Berndt et al., 2006).

Innovation in task performance involves two aspects – problem discovery and problem solving (von Hippel and Tyre, 1995; Hyysalo, 2006). Performance with respect to both of these aspects may be affected by organizational issues, such as management structure or processes for

resource allocation, which may either support or interfere with second-order problem solving and innovation (Tucker et al., 2002; Adler and Clark, 1991). In this paper, we propose that such innovation is also affected by the organization of routine work itself.

The positive relationship between specialization and productivity has been previously shown at the organizational, divisional and plant levels (Brush and Karnani, 1996; Huckman and Zinner, 2007; Vokurka and Davis, 2000). Studies on individual level productivity at the team level have also been conducted. Narayanan et al. (2009) use team level data in a software maintenance firm to show that specialization improves individual productivity but exposure to variety at the workgroup level has non-linear effects – decreasing productivity at very high levels. Similarly in an experimental setting, Schilling et al. (2003) find that specialization improves individuals' outcomes but only as much as doing an unrelated task. Doing a different but related task, on the other hand, improves outcomes much more than specialization (Schilling et al. 2003).

A study by Boh et al. (2007) on the other hand, finds that while specialization increased productivity at the individual level, diverse but related experience is more important for team and organizational performance in software maintenance projects. Furthermore, Staats and Gino (forthcoming) show that the effect of specialization and variety may be different in the short and the long run – in the short run, specialization improves productivity but in the long run, variety is more important because workers learn how to learn better (Staats and Gino, forthcoming).

The importance of specialization and variation in work on performance has also been explored in the health care industry. A study by Diwas and Staats (forthcoming) looks at the importance of focal and related experience on outcomes in minimally invasive cardiac surgery procedures to find that focal experience has a greater effect on surgeon performance than related

experience. Similarly, “procedural specialization” -- i.e. the type of procedures that a surgeon performs rather than educational specialization – has been shown to be correlated with lower patient mortality rates among general surgeons (Hall et al., 2009).

The impact of diversity on innovative performance has been studied at the organizational and team levels. At the organizational level the scope of R&D programs has been shown to positively impact innovative outcomes (Cockburn and Henderson, 2001). Theoretical work on team level diversity has found that a randomly chosen group of diverse problem solvers outperforms a select group of high ability problem solvers precisely because high ability problems solvers are similar to each other (Hong and Page, 2004). In reality however, teams with diverse members are prone to conflicts and as a result the effect of diversity on performance depends on team process variables (Pelled et al., 1999; Bunderson and Sutcliffe, 2002; Rico et al., 2007; Bettenhausen, 1991; Williams and O’Reilly, 1998). For example, Ancona and Caldwell (1992) found that diversity had a direct negative effect on dimensions of innovation in new product development teams. The indirect effect of team diversity however, through external communication, was positive.

Diversity of individual experience has also been studied within the context of the team with mixed results. In a non-innovative context, intrapersonal functional diversity in top management teams has been shown to have a positive effect on information sharing and unit performance (Bunderson and Sutcliffe, 2002). However, in contrast to these positive findings, Astebro and Thompson (2011) show that people with a more varied experience in the general population select to be entrepreneurs but conditional on being entrepreneurs, those with more varied experience earn less (Astebro and Thompson, 2011). In terms of innovation, research has shown that the share of “multi-knowledge” individuals (those with diverse functional

background) in new product development teams has an indirect positive effect on product innovativeness through information sharing (Park et al, 2009).

The impact of focus and variety in routine work on innovative outcomes at the individual level, however, has not been studied. Levitt and March (1988) propose that at the organizational level experience and improvement on an inferior task can result in a "competency trap" which can prevent superior procedures from being searched for and adopted (Levitt and March, 1988). Individual level theories have not been proposed or tested and it is unclear whether the focus of routine work impacts innovation at all. The direction of such a relationship also needs to be explored.

There are a few ways in which the diversity of routine work can impact innovation. A diverse set of experiences can enable a worker to identify problems in one area of experience by comparing it to another. For example, a surgeon who is performing laparoscopic surgery using specific instruments for one area of the body and a different set of instruments for another may notice that the difficulty of operating varies by the type of instrument used. Had she not varied her work, she would not have thought about applying one set of instruments in a different setting to improve suboptimal outcomes. With varied experience she can suggest improvements to the tools (see Gauderer, 2009 p. 17 and Riskin et al., 2006 for multiple examples).

Similarly, surgeons who do a variety of procedures may be better able to apply techniques learned and improved in one task of a particular procedure to problems encountered in a different procedure. If new problems are encountered in the process of transfer of the technique from one procedure to the other, further improvements may be possible. In fact, analogical thinking has been shown to impact problem identification, problem solution and innovation (Christensen and Schunn, 2007; Kalogerakis et al., 2010).

A diverse set of experiences can impact innovation also by allowing workers to combine different ideas or techniques to create new and potentially more efficient ones. In fact recombination has been described as a way to create innovation in numerous previous studies (Schumpeter, 1939; Henderson and Clark, 1990; Nelson and Winter, 1982; Fleming, 2001).

It is possible, however, that focus in a certain area, rather than experience with a variety of tasks may be more important for innovation. Forgetting is a documented phenomenon both at the individual and at the organizational level (Johnson and Hasher, 1987; Newell and Simon, 1972; Argote and Epple, 1990; Argote, Beckman and Epple, 1990). Forgetting at the individual level then would imply that people are less likely to use and build on the knowledge that they acquire from a task if that task is followed by a much different one. In fact, Simon (1990) estimates that approximately seven chunks of information can be stored in an individual's short term memory. Repeated exposure is then needed for patterns and rare deviations from patterns to be recognized (Narayanan et al. 2009). Focus, rather than varied experience, would then allow for such problem recognition and innovative solution.

Because there is no previous research that would tell us in what direction diversity of experience may influence innovation, we propose the following hypothesis:

Hypothesis: ***Diversity of routine clinical work (exploitation) will influence the quantity and quality of innovative output (exploration) by cardiac surgeons.***

To test this hypothesis we use a dataset of cardiac surgeons in NY that work at major teaching hospitals. We have diagnosis and procedure information on every patient they saw between 1994 and 2004 as well as information on their academic publications. Cardiac surgeons at major teaching hospitals have dual roles - their daily clinical work measured by the procedures

that they perform represents their exploitation activity and their academic output represents a report of their innovative ideas and discoveries – a reflection of exploration.

Furthermore, surgeons have a long history of innovating both in technique and in devices. Chatterji et al. (2008) find that 20% of medical device patents have at least one physician as an inventor with surgeons constituting 20% of those inventors. Furthermore, patents by physicians are more important and more general as seen from analysis of patent citations received (Chatterji et al., 2008). In cardiac surgery, for example, one of the most important medical devices – the cardio-pulmonary bypass machine – was created by a surgeon, John H. Gibbon who tested his first prototype in 1937.

Surgeons' expertise in devising new medical devices is also important because of the tacitness of the information that they possess with regard to the techniques that they employ. It is impossible to encode information regarding the movement of one's hand close to a beating heart for example, or the way an instrument glides in relation to a blood vessel. As a result, problem identification almost invariably occurs by the surgeon and problem solution is often done iteratively with surgeons and firms working together (von Hippel, 1994; Chatterji et al., 2008).

Besides medical devices however, surgeons routinely innovate on techniques as well. Examples are many and varied and span the history of surgery itself (Mehta, 2009; Riskin et al., 2006; Gauderer, 2009). Modifications by fellow surgeons eventually make new techniques safe and mainstream (Starr, 2010).

We try to understand how the diversity of clinical experience influences the quantity and quality of surgeons' exploration as proxied by academic publications. This setting is particularly good for testing our hypotheses because surgeons' practical experience consists of their clinical work while their academic output also reflects and is based on that clinical work. Practicing

surgeons rarely conduct lab research unrelated to their clinical work. In addition, their academic work is a good proxy for innovation since peer reviewed academic journals generally publish work containing new ideas or findings. Furthermore, surgeons vet new ideas and techniques in the community through these publications.

What is important for our setting is that surgeon innovations are reflected in academic articles that can be observed and whose quality can be proxied using widely accepted measures. Articles range from problem descriptions, longitudinal studies that identify correlations to descriptions of new techniques and clinical studies using those techniques. As a result, this provides us with the perfect setting to test our hypothesis.

3.4 Description of Data

3.4.1 Data on Surgeons

The main data for our paper comes from New York State's Statewide Planning and Research Cooperative System (SPARCS) and includes all inpatient hospital stays in New York State between 1994 and 2004.[1] Each patient-stay is a unique observation and contains up to 15 diagnoses and up to 15 procedures as well as separate attending physician and surgeon license numbers which we use to link to surgeons' academic profiles.

3.4.1.1 Surgeon Selection

For our paper we use a sample of 162 surgeons who have worked at least one year at a major teaching hospital in NY and have performed at least 25 Coronary Artery Bypass Graft (CABG) or valve procedures in at least one of the 11 years they are in our dataset. CABG and valve are the routine operations that cardiac surgeons perform and most every cardiac surgeon

[1] For more detailed information see http://www.health.ny.gov/statistics/sparcs/

performs them. To be in our sample, these surgeons also had to be listed as thoracic, cardiothoracic or cardiovascular surgeons in one of three places: 1) the Cardiothoracic Surgeon Network website (www.ctsnet.org)[2] 2) the "Coronary Artery Bypass Surgery in NY State" reports for 1994-2004[3] or 3) the www.nydoctorprofile.com[4] website which provides information on a physician's education and specialty. They also had to 4) have performed at least one surgery in a major teaching hospital during this time period.

Restrictions 1) and 2) we put in place because of intricacies of the data. The data lists a doctor and a surgeon responsible for the primary surgery on each patient observation. However, for people who had some other procedure together with a CABG or valve procedure, such as trauma surgery, the surgeon listed on the patient observation could be different from the cardiac surgeon. So based on just the first selection criterion we ended up getting about 13 other doctors who were cardiologists, trauma surgeons, emergency room (ER) doctors in our sample. Since we are not interested in following their clinical experience as it is not comparable to the rest of our sample, we decided to exclude them.[5]

The last restriction is put in place because we are interested in article publications and not all surgeons publish. Major teaching hospitals are our proxy for incentive to publish and come from a list by the American Hospitals Directory. They are certified by the Council of Teaching Hospitals (COTH) and have a certain number of resident spots to qualify as a major teaching

[2] http://www.ctsnet.org/sections/members/surgeons/ accessed in 2008.

[3] http://www.health.ny.gov/statistics/diseases/cardiovascular/ accessed in 2007 and 2008.

[4] http://www.nydoctorprofile.com/search_parameters.jsp accessed in 2008.

[5] Based on the fact that only 13 surgeons who were not cardiac surgeons passed our first test over 11 years of data, we believe that the omission of the cardiac surgeon as the surgeon is rare in the data and does not affect our cardiac surgeons' clinical experience significantly but we don't have other ways in which to check that claim.

hospital.[6] Furthermore, surgeons who never publish are excluded from our sample because we rely on a panel dataset and their outcome variable does not change.

We also exclude those surgeon-year observations in which a surgeon has fewer than 25 total patients. Those are, with six exceptions, years in which the surgeon is first or last observed and we are not sure that the surgeon practiced in New York State for the whole year. For our analysis we are left with a sample to 1307 surgeon-year observations – an unbalanced panel of 162 surgeons over 11 years. Similarly, we also report results with samples that include only surgeons who have fewer than 50% of their patients receiving non-cardiac procedures throughout their career. That restriction was put in place because we were worried that surgeons who were doing predominantly other thoracic surgery such as lung procedures were not comparable to the rest of the sample.

3.4.1.2 Clinical Focus

Once we select our sample of surgeons we proceed to select each patient that they have treated. Detailed patient observations are then combined to create a composite measure of the diversity of different types of procedures that a surgeon performs in a certain year. Each patient observation is a stay at a hospital (we removed duplicate observations for ancillary service use) and contains a number of detailed variables on patient demographics, attending physician, surgeon, hospital, payment as well as up to 15 diagnoses and 15 procedures performed. For each patient we use the procedures reported using standard International Classification of Diseases Revision 9, Clinical Modification (hereafter ICD-9) codes and sum those over all patients of every surgeon for every year to arrive at surgeon-year level observations for our models.

[6] http://www.ahd.com/definitions/prof_teach.html accessed in July 2008.

Particularly important for our purposes is that the license numbers are reported for the attending physician and the surgeon that treat the patient. We use the surgeon license number to find additional surgeon specific information such as education, year of graduation and licensing and, most importantly, articles published. The data also contains information on the hospital at which the surgery was performed, which helps us determine whether it is a major teaching hospital and whether the physician has an incentive to publish or not.

We use the 15 procedure variables to understand the make of the physician's clinical work. To do this we need to first determine what procedures a cardiac surgeon would perform, how they are similar to each other and can be grouped together in categories. The reason why we group ICD-9 procedures into different categories is that ICD-9 codes are very detailed and two different ICD-9 codes do not necessarily reflect very different tasks that could potentially result in variation in learning. For example, a surgeon may perform a CABG where he only bypasses one artery (ICD-9 – 3611) or one where he bypasses two arteries (ICD – 3612). While the second procedure is more complex, it is not very different from the first procedure if split into its constituent tasks. We create measures of focus based on these categories and include those in our models. There are a large number of ways in which these categories can be created and below we report three such groupings.

There are no previously accepted standards for determining what constitutes a cardiac surgeon's work and cardiac surgeons are also trained in general surgery so could potentially perform other surgeries as well. In addition, while the majority of cardiac surgeons' work is indeed on the heart, many cardiac surgeons are trained and often perform surgeries on the lung or the other thoracic organs as well. To be able to compare them we need to look at the cardiac

surgeries that they perform and then compare them on those particular sets of tasks. Below we report our three different measures of cardiac surgeons' clinical focus.

3.4.1.2.1 Task Based Focus Measure

The first measure, which we call *Task Based Focus* measure is based on our own understanding of cardiac surgeons' work and how the procedures they do may differ based on the tasks required to complete each. That measure divides the procedures into six categories – 1) CABG, 2) Heart Valve, 3) Heart Muscle (other than heart valve), 4) Heart Vessel (other than CABG), 5) Pacemaker/Defibrillator and 6) Heart Assist Device/Transplant procedures. While CABGs are part of the heart vessel surgeries we decided to separate them from the rest of the Heart Vessel procedures because of the large volume of GABG surgeries compared to other cardiac surgeries. For the same reason we separated Heart Valve procedures from Heart Muscle procedures. We also presume that routine surgeries such as CABG and Heart Valve have been more standardized and innovation may be less likely in them. Below we describe the procedures in each of the categories. A list of the ICD-9 procedure codes that we used in each category is included in Appendix 3.A.

1) **CABG -** involves bypassing a section of a coronary artery with a graft from a different blood vessel from another part of the patient's body

2) **Heart Valve** – repair or replacement of a heart valve

3) **Pacemaker/Defibrillator insertion -** simpler procedures than the rest that do not necessarily require opening the chest wall and can be performed by general surgeons. In our case, they are only included if performed by a cardiac surgeon. Often they are auxiliary to other surgeries.

4) **Heart Assist Systems /Transplant** – includes insertion of left ventricular assist devices, other complex artificial systems that assist the heart muscle in doing its job in severely ill patients with congestive heart failure. These devices normally assist patients who wait for transplants. Because of the low volume of heart transplants, we include them in this category.

5) **Heart Muscle** – these include procedures on the ventricles, heart atrium, myocardium, the pericardium and other muscular structures of the heart.

6) **Heart Vessel**– treatments of heart and thoracic vessel aneurysms and revascularization other than CABG. Includes procedures performed on the thoracic aorta.

7) **Non-cardiac Procedures -** includes lung cancer surgeries, other thoracic procedures such as on the diaphragm, trauma procedures, diagnostic procedures such as cardiac catheterization and anything else that does not have an ICD-9 procedure code that is related to open heart cardiac surgery.

An observation (i.e. patient stay) was classified as non-cardiac if no cardiac procedure was reported from the ones in categories 1) through 6). We previously did have a separate “lung” category but our understanding at the moment of lung procedures is not sufficient to help us classify them into different more detailed categories within the lung surgeries group. In further work on this paper, we are planning on distributing a survey to a few surgeons that will ask them to independently assign lung procedures to different categories.

The table below shows a frequency distribution of the six categories along with non-cardiac procedures performed by our 162 surgeons over 11 years in our dataset.

Table 3.1 Frequency distribution of cardiac procedures across Task Based Focus categories

Category	**Number of Procedures in Each Category (all surgeons, years)**
Coronary Artery Bypass Graft (CABG) Procedures	119,879
Valve Procedures	38,821
Heart and Thoracic Vessel Procedures (no CABG) - HTV	6,723
Heart Muscle Procedures Other Than Valve	21,791
Heart Assist and Heart Transplant Procedures	9,973
Pacemaker/Defibrillator Procedures	18,287
Total Cardiac Procedures	**215,474**
Total Cardiac Patients	**175,484**
Non-cardiac Patients	41,137
Total Patients	**216,621**

Note that the total number of patients is different from the sum of the total procedures. This is because if a patient gets procedures from two or more different categories, we count each of them separately. However, if a patient gets two procedures from the same category, we don't count them as two separate procedures. For example, if a person received a pacemaker and a defibrillator, then that would be counted as only one procedure in the Pacemaker/Defibrillator category. If she, however, received a valve replacement and a pacemaker, then we would count these procedures in each of their respective categories. From the data above, we see that our cardiac patients received on average cardiac procedures from 1.23 of the above categories.

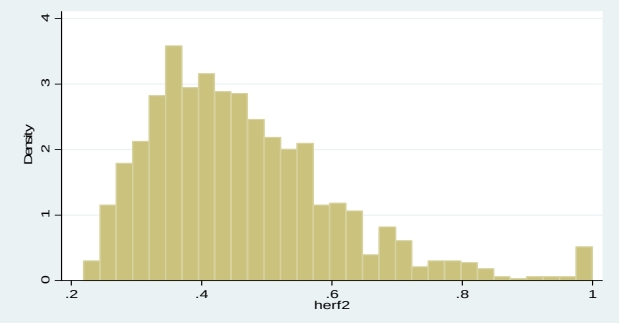
To measure the diversity of a physician's clinical work we create a Herfindahl-Hirschman (HHI) index based on the share of the different cardiac procedure categories above in her total cardiac procedure count i.e. we sum the squares of the share of Heart Assist/Transplant procedures, the square of the share of CABG procedures and so on with all six cardiac procedure categories. In creating the shares, we use the total number of cardiac Task Based Focus category procedures. In this measure we exclude the Non-cardiac category. A large Non-cardiac category share will inflate our HHI while possibly containing a very diverse set of procedures itself. In our

models we include the share of patients with no cardiac procedures along with the cardiac HHI. Below is the formula for our Herfindahl-Hirschman measure:

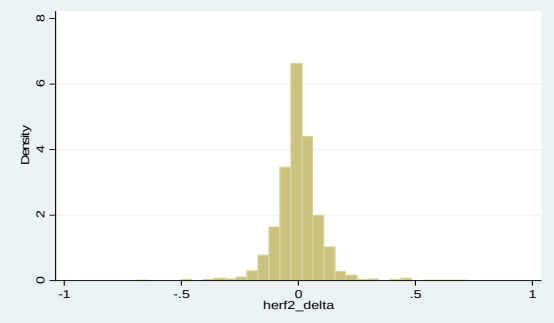
$$HHI = \left(\frac{c_{itk}}{S_{tk}}\right)^2,$$

where c_{itk} is the number of procedures in cardiac category i in year t for surgeon k and S_{tk} is the sum of procedures across all cardiac categories in the same year for surgeon k. Below are graphs of the distribution of the HHI index based on the Task Based Focus Categories.

Graph 3.1a Frequency Distribution of the HHI index based on the Task Based Focus categories across all 162 surgeons over 1994-2004



Graph 3.1b Frequency distribution of the Yearly Change in the HHI Index within Surgeons over all surgeons, all years



3.4.1.2.3 CABG/Valve Based Focus Measure (CVL Focus)

Because CABG and Heart Valve procedures are the main ones that the surgeons in our data perform, together constituting 74% of all cardiac procedures, we decided to create a measure that separates the categories from the Task Based Focus measure into larger groups centered on CABG and Heart Valve procedures. In this CVL measure we have four cardiac surgery categories – “CABG”, “Heart Valve”, “CABG and Valve” and— “No CABG or Valve”. Non-cardiac surgeries are excluded. These categories are mutually exclusive by patient – i.e. a patient cannot have a procedure in more than one of these categories. This is different from the previous measure where a patient could have a procedure in all six different categories

simultaneously – for example in CABG, Heart Valve and Pacemaker/Defibrillator. For that reason, in calculating the share of a specific category in the HHI index the numerator was the number of cardiac procedures in a certain category performed by the surgeon in the specific year and the denominator was the total number of cardiac procedures in that year by that same surgeon. When we calculate shares in the CVL categories the numerator is the number of patients who had a procedure in the specific CVL category performed by a specific surgeon and the denominator will be the total number of patients that had a cardiac procedure in that year by that surgeon. Below is a description of the categories in the CVL focus measure:

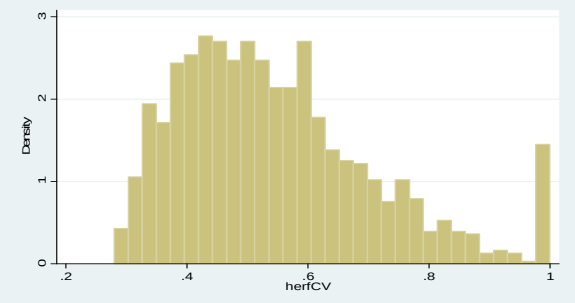
1) **CABG -** the patient has received at least one procedure in the CABG Task Based Focus (TBF) category but no procedure from the Heart Valve TBF category. Patient may have received cardiac procedures from the other TBF categories.
2) **Valve** – the patient has received at least one Heart Valve procedure but no CABG procedures from the TBF categories. Patient may have received more cardiac procedures from the other TBF categories.
3) **CABG and Valve** – the patient has received both a CABG and a Heart Valve procedure from the TBF categories. Patient may have received cardiac procedures from the other TBF categories.
4) **No CABG or Valve Cardiac** – the patient has received at least one cardiac procedure from at least one TBF category but no procedures in the CABG or Heart Valve TBF categories.

Descriptive statistics and graphs of the CVL based categories are below.

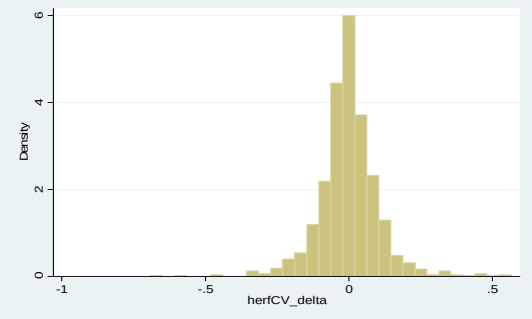
Table 3.2: Frequency distribution of cardiac procedures across CABG/Valve Focus Measure categories

CVL Category	Number of Procedures in Category
Patients with CABG	106,397
Patients with Valve	25,339
Patients with Both CABG and Valve	13,482
Patients with Cardiac Procedures but No CABG or Valve	30,266
Non-Cardiac-Surgery Patients	41,137
Total Patients	**216,621**

Graph 3.2a: Frequency distribution of the HHI index based on the CABG/Valve Based Focus categories across all 162 surgeons over 1994-2004



Graph 3.2b: Frequency distribution of the Yearly Change in the CVL HHI Index within Surgeons over all surgeons, all years



3.4.1.2.3 Clinical Classification Software (CCS) Based Focus

Our third focus measure is based on a categorization of diseases and procedures designed by the Agency for Healthcare Research (AHRQ) as part of the Healthcare Cost and Utilization Project (HCUP). According to its designers, the CCS classifies diseases and procedures "into a smaller number of clinically meaningful categories" based on the ICD-9 codes. [7]

The CCS system has three levels of detail and for our purposes we use section 7 – Cardiovascular Procedures at the second level of detail. We further select only operating room (OR) - based procedures from the second level and exclude non-cardiac vascular procedures

[7] http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp accessed December 25, 2011

such as 7.14 - *Other vascular bypass and shunt; not heart* or procedures that would be performed by cardiologists and not cardiac surgeons such as 7.5 *Diagnostic cardiac catheterization; coronary arteriography* or *Percutaneous trans luminal coronary angioplasty (PTCA).* Appendices 3.B1 and 3.B2 describe section 7 of the CCS classification at the different levels of detail. For our purposes we retained five categories at the second level of section 7 of the CCS classification described below:

7. 1 Heart Valve procedures

7.2 Coronary artery bypass graft (CABG)

7.6 Insertion; revision; replacement; removal of cardiac pacemaker or cardioverter/defibrillator (Pacemaker/Defibrillator)

7.7 Other OR Heart procedures

7.10 Aortic resection; replacement or anastomosis

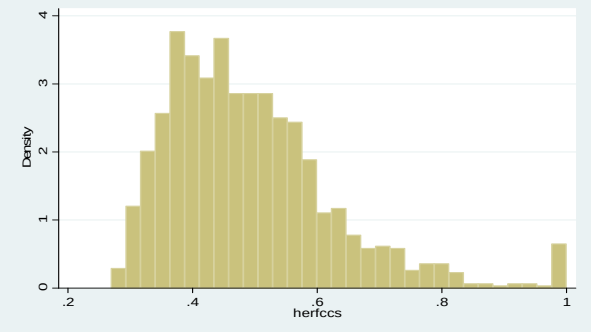
Section 7.10 –*Aortic resection; replacement or anastomosis* was absorbed in section 7.2 - CABG - as the procedures are similar and there are very few observations in 7.10 in our dataset. Similarly, cardiac transplant procedures are not part of section 7 of the CCS at all and there are again very few such procedures so they were added to the "Other OR Heart procedures" category

The CCS categories are comparable to the Task Based Focus Measure categories but differ in certain significant ways. Many procedures in the TBF based Heart Vessel category are classified in the CABG category in the CCS classification even though they are not technically CABG. Similarly, a fair number of the TBF based Heart Muscle procedures are in the Heart Valve category according to the CCS classification. However, a large number of the TBF based Heart Vessel and Heart Muscle categories are also classified in the "Other OR heart procedures" category.

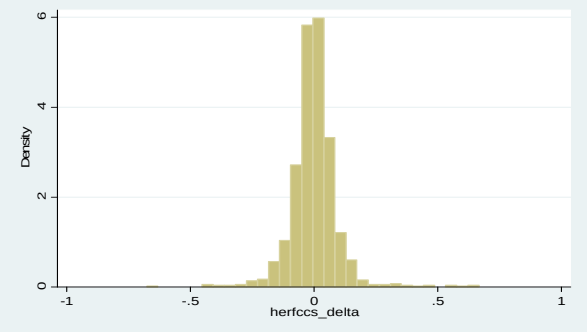
Table 3.3 Frequency distribution of cardiac procedures across CCS Focus Measure categories

CCS Category	Number of Procedures in Each CCS Category
CCS Valve Procedures	38,857
CCS CABG Procedures	122,233
CCS Other OR procedures	31,671
CCS Pacemaker/Defibrillator Procedures	17,764
CCS Total Cardiac Procedures	**210,525**
CCS Total Cardiac Patients	**173,716**
CCS Non-cardiac	42,906
Total Patients	**216,621**

Graph 3.3a Distribution of the HHI index based on the CCS classification based focus categories across all 162 surgeons over 1994-2004



Graph 3.3b Distribution of the Yearly Change in the CCS HHI Index within surgeons over all surgeons, all years 1994-2004



3.4.1.3 Share of total procedures based measures

Another way to understand the clinical work of cardiac surgeons is to examine the influence of specific types of procedures they perform on their innovative outcome. Some procedures are significantly different from others in terms of the tasks needed to complete them and their complexity. Certain areas of academic inquiry and surgical innovation may be newer and open to more contributions because standards have not been established in the community. For example, there may be more opportunity to innovate in treating congenital heart diseases that

tend to be more complex and rarer than CABG, which due to the high incidence of cardiovascular disease has become a more routine procedure. Furthermore, the effect of specific procedures on articles related to that procedure or to other procedures could indicate experiential and innovative spillovers. For that reason, we run a number of models in which we look at how the change in the share of a certain category of procedures in a surgeon's work impacts the number and quality of articles that she publishes.

3.4.2 Data on Academic Articles

We measure academic output by the number of articles that a surgeon has published in peer-reviewed journals. We exclude editorial materials, meeting abstracts, bibliographies, letters, review articles or notes which do not necessarily report innovative outcomes and are not always peer reviewed. Our interest is in innovative output which is often measured by patents but few of our surgeons have actually filed patents. In addition, patents are good indicators for innovations involving medical devices but not for innovations in technique or procedure as the latter kind of patents are hard to enforce. Academic journal articles are by definition innovative because they put forward new insights that physicians glean from their clinical or lab work (Schroeder et al., 1989; Merton, 1957; Stephan, 1996).

We obtain our data from the ISI Web of Science citation index. The Web of Science (WOS) database provides detailed information about the articles. It includes the title of the article, all the authors and their addresses as supplied by them, subject and separate keywords reported by the authors or marked by WOS. In addition to the source and year of publication, the number of cites each article has received from publication till the current time (the date the data is gathered by me in this case) is also reported.

We search for surgeon-authors in this database by their last name and a combination of their first and, if available, middle initial, as well as at least one address in the state of NY on the article. To make certain that the article belongs to the particular physician we compare each article's address to the physician's hospital at the time it was published and the year prior to publication. We have the name of the hospitals at which they practiced from the SPARCS dataset and use the Cardiothoracic Surgeons' Network website at www.ctsnet.org as well as the NY State CABG reports for further checks of their hospital of practice during specific years. We also use additional checks such as the topic of the surgeon's article and the availability of other physicians with the same last name and initials in the state. For example, there are two physicians named Ko Wilson (no middle initial) in NY. One is an ophthalmologist and the other is a cardiac surgeon and they write on two very different topics. Another example is of a father and a son with the same last name and two initials that match (JJ Rose). One is a plastic surgeon and the other is a cardiac surgeon. At some point in their career they also work at the same university hospital. A detailed review of the articles based on the points above lets us make the correct article attribution.

Our article data spans the years from 1994 through 2004. In our models we lag our focus data by a year in comparison to our article data based on the assumption that articles conceived in a certain year will be published in the next one. This is reasonable given the fact that the turnaround time for medical journal articles is about 7 months[8], much shorter than that for economics and management journals.

[8] http://www.californiahealthline.org/articles/1999/10/26/MEDICAL-JOURNALS--More-Troubles-for-NEJM-JAMA.aspx?archive=1

In our models we use citations as an indicator of innovativeness and quality of a surgeon's research (Adams and Griliches, 1996; Azoulay, P, 2002).[9] They are, however, a very imperfect measure of quality. One of the biggest criticisms has been that citations do not necessarily reflect building on knowledge. For example, citations may reflect a scientific fad that is not based on the merit of ideas. Self-citations pose additional problems. Furthermore, citation propensity likely varies by field and subfield. Additionally, articles that report on more basic science have been found to be cited more than articles on more applied and narrower topics which can be just as valuable from society's point of view (Lindsey, 1989; Cockburn et al., 1999). However, we believe that in our case articles are a good proxy for the quality and innovativeness of published research as they are in the same field and most likely occupy a similar space in the basic/applied research continuum. Also, they have a strong history of use in economics and management literature (Cockburn and Henderson, 1998; Azoulay, 2002).

From the article data, we create two measures of a physician's academic output. The first one is a yearly variable that measures the number of articles a surgeon has written in each year. To control for the fact that many medical articles have more than one author and that more authors imply that an article has received less input from each specific author, each author gets credit for only 1/n-th of an article that has n authors. This is based on the fact that in the sciences and to a large extent in medicine as well, the names of all the lab or departmental collaborators will be on the article with the first author being the person in charge of the experiments and the write-up of the results and the last author being the PI of the lab.

---

[9] Citation statistics have been used to indicate quality not only for research purposes but also by governments and institutions to determine research spending output. http://www.theaustralian.news.com.au/story/0,25197,23990703-12332,00.html

Our second outcome measure uses article cites to quantify the quality and innovativeness of each physician's academic output. This measure counts the citations an article has received per year since its publication and is also divided by the number of authors. It is then summed over all the articles the surgeon has published in a given year to yield a cite-weighted article count. This measure does not control for the number of articles a surgeon has written in a given year. For example, a surgeon who has written 5 articles that each received 2 citations per year will have the same value as a surgeon who wrote 2 articles that each received 5 citations per year. In that way, this is a measure of the quality of the overall research of a surgeon per year. This measure also helps compare research quality among people who publish on many different topics and those that publish only on a few. This means that a surgeon who only performs CABGs and may only write on the topic of CABG but write high quality articles will be the same on this measure as someone who does many different procedures in different categories and writes less important articles on each of the different topics – i.e. CABG, valve, heart transplant etc. This controls for the fact that surgeons who work in different areas have the opportunity to publish in many different areas as well.

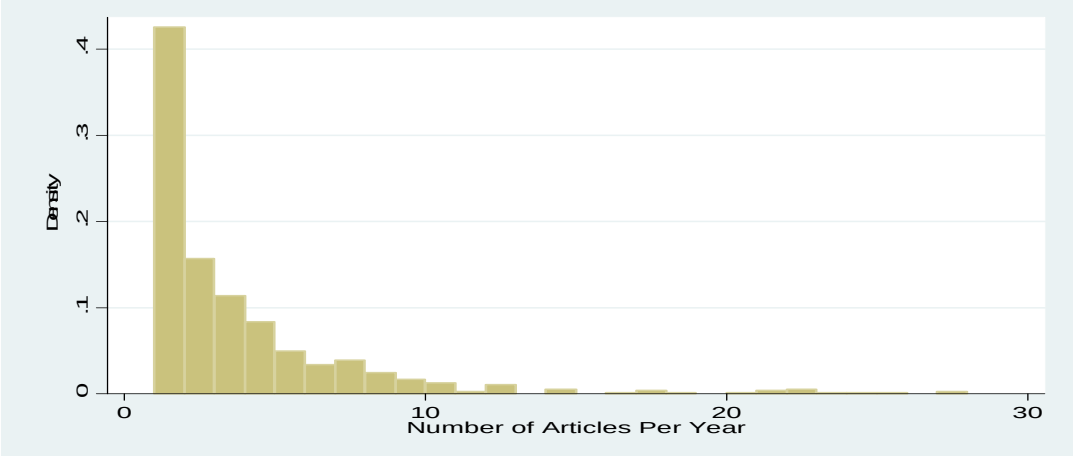
Due to data constraints, we are only able to get information for number of citations as of June 2008 (when we finalized the data gathering process). We then use the years passed since the publication of the article to find the average citations per year. This assumes that citations are uniformly distributed over the years. There is no data to support or disprove this assumption. It is possible, for example, that articles reach a peak yearly citation rate at some time after they are published possibly at 5 years after publication. For articles that have been out for less than 5 years then, we may be under-representing their quality. This is a shortcoming of our data.

For some of our analyses we also divide the articles by topic. We use only two topics – valve and CABG since the rest are hard to classify into meaningful but small number of categories. We use a number of keywords in the title, subject, keyword or abstract to classify articles into these two categories. Below are some summary statistics about our articles and a graph of their frequency distribution:

Table 3.4 Article Publications Data

Category	**1994-2004**
Total number of articles by authors in our dataset	2486
Total CABG articles	396
Total Valve Articles	349
Total First Author articles by surgeons in our data	274
Lowest number of articles per year per surgeon	0
Highest number of articles per year per surgeon [not divided by number of authors per article]	28

Graph 3.4 Frequency Distribution of Article Publications per Year per Surgeon



3.4.2 Data for Control Variables

In our models we also include a few procedure related and surgeon related control variables that we think may be correlated with both surgeon article output and their clinical

focus. As mentioned before, we have a panel dataset and our dependent variable is the yearly change in (cite-weighted) article publications by a specific surgeon and our independent variable of interest is the change in the surgeon's clinical focus. However, over time, as surgeons advance in their career and are more sought after by patients they may have more influence over the procedures that they elect to perform and hence over their clinical focus. For that reason we control for the surgeon's career age. This variable is defined as the time from the surgeon's graduation from medical school and was obtained from the New York State Department of Education Professional Licensing Office web site.[10] Data were checked and complemented by surgeon self-reported graduation dates from the Cardiothoracic Surgery Network (CTS) website.[11]

There are certain problems with measuring career age from time of medical school graduation rather than the starting date of practicing as a cardiothoracic surgeon. Some may decide to become cardiac surgeons and start their training right after medical school while others may have practiced as general surgeons for some years before becoming cardiothoracic surgeons. In terms of career age and influence on patient mix then, they would not be similar in our dataset even if they graduated at the same time from medical school. Unfortunately, this is the best data that is available. We considered professional licensing data but that licensing is often received immediately after medical school graduation, not at the start of a cardiothoracic career. There is also an additional downside to using licensing data because it is by state and surgeons who practiced in a different state before coming to NY and received they NY State license later would be misrepresented. Our measures of career age based on time from medical

[10] http://www.op.nysed.gov/opsearches.htm accessed during 2008

[11] www.ctsnet.org accessed during 2008

school graduation and professional MD licensing are also highly correlated – with a correlation coefficient of 0.87.

Another variable that we suspect is correlated with both the mix of procedures that a surgeon performs and her academic articles is the complexity of the types of procedures that she performs. A surgeon who performs more complex surgeries may also have an opportunity to learn and innovate more. We measure complexity by the number of procedures from different task based focus categories that the surgeon performs on each patient – a patient that receives both a Heart Valve and a CABG procedure would be a more complex case than a patient who just received a CABG procedure. Note, however, that multiple procedures from the same focus category on the same patient will be only counted as one procedure. To get our surgeon year value of complexity we count the procedures that each patient received from our different categories and average them over the total number of patients for each surgeon. Our complexity variable has a range between 1 and 1.84.

This complexity measure is re-calculated for our CCS focus measure because the main categories are different from the Task Based Focus measure. For example, many cardiac vessel procedures that would be in the Heart Vessel category in our Task Based Focus measure are included in the CABG category. As a result, the complexity measure is different for patients who would have had procedures in two different Task Based Focus categories but now have procedures only in one CCS category. The range of the CCS complexity measure is 1 to 1.68 due to the smaller number of different categories.

Our CABG/Valve based focus measure is designed in such a way that each patient belongs to one of four categories, rather than procedures belonging to categories. The four categories are defined based on whether the patient received: “CABG”, “Heart Valve”, “CABG

and Valve" and "Cardiac, but No CABG or Valve" procedures. As a result, complexity cannot be defined using the four patient categories above. We thus use the Task Based Focus categories to create a complexity measure and include it in addition to the CABG/Valve based focus measure in these models. Our results remain the same in direction and statistical significance if we do not include a complexity measure at all in the CABG/Valve focus specification.

3.5 Models and Results

In all our models, we use Poisson Quasi Maximum Likelihood estimation with surgeon fixed effects and clustered standard errors by surgeon. The Poisson QML is a count model with estimates that are consistent without the assumption of a particular underlying distribution of our variables. Fixed effects OLS provides very similar results which have not been reported here. Because it is a fixed effects model, the XTPQML model drops all surgeons who have no article publications or cites in the respective models throughout their career.

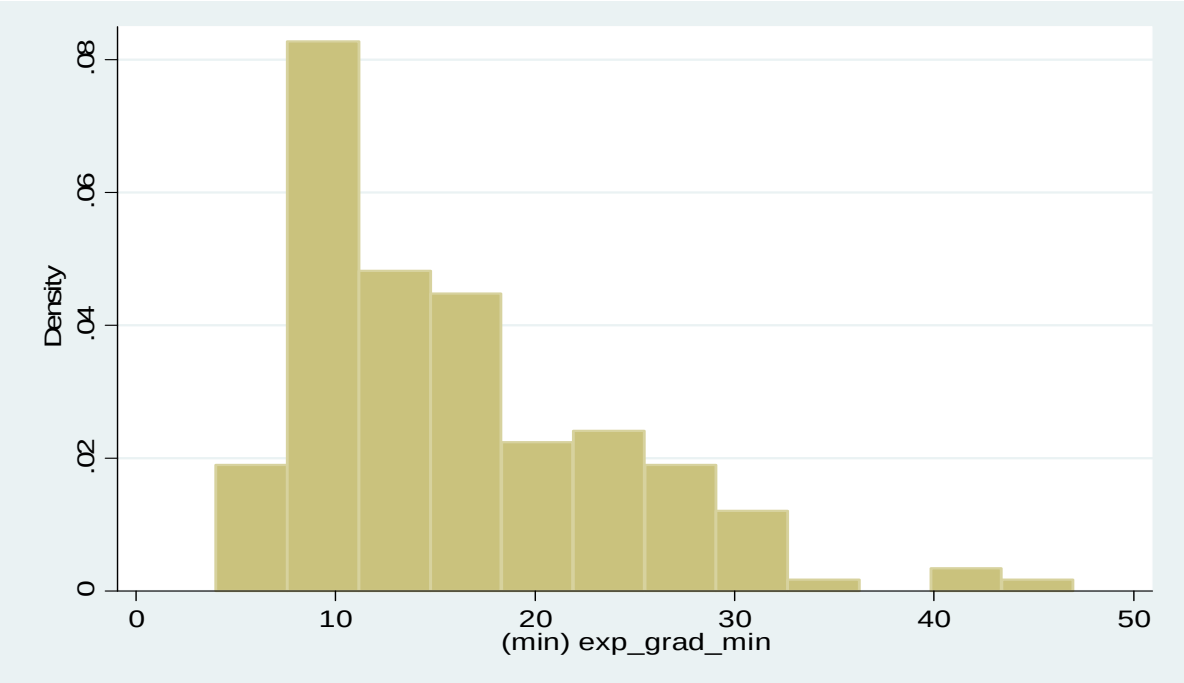
For our identification strategy, we use within surgeon variation of clinical and academic measures across different years. Our goal is to avoid potential omitted variable bias such as surgeon innate ability or curiosity which may be responsible both for someone's higher diversity or focus of clinical work and higher publication output. We also include a yearly trend variable by introducing surgeon's professional experience which measures the time elapsed since a surgeon graduated from medical school and changes every year as one more year is added to her experience.

Our main results are reported in Tables 3.5a-3.5c. Each of the tables contains results with two different dependent variables: all cite-weighted articles per year per surgeon in the first three columns and the raw number of all articles per year per surgeon in the last three columns. The main variable of interest – our HHI based focus measure, differs in each table based on the

categories used to construct it. In Table 3.5a – we use our own knowledge of cardiac surgery to classify each procedure into one of the Task Based Focus cardiac categories – CABG, Valve, Heart Muscle, Heart Vessel, Heart Assist/Transplant and Pacemaker/Defibrillator. We then use the share of each of these categories in a surgeon's total volume of heart procedures to construct a Herfindahl-Hirschmann index (HHI) which we include as our main independent variable. In Table 3.5b the HHI is based on dividing the above categories into coarser groups where each cardiac patient belongs to one of four categories - CABG, Valve, CABG and Valve, Other Cardiac patient. The last HHI in Table 5c is created by using the CCS categories with small modifications explained in our data section.

In each of the tables, the first and fourth columns include results from the whole sample. The rest of the models are run on subsamples of early-career and late-career surgeons. The division between the early and late career surgeons is made by taking the years since graduation variable for each surgeon at the time when he enters the dataset. The distribution is skewed with a mean of 15.4 and a median of 14 years. We select the median as our division point. Surgeons with experience of less than 14 years at time of entry into the dataset are classified as early-career surgeons and those with 14 or more are classified as late-career surgeons. Below is a histogram of the years since graduation variable for each of the 162 surgeons in our dataset. It is measured at the time when we first start observing them. Tables 3.5a, 3.5b and 3.5c with our results are on the following page.

Graph 3.5 Histogram of "Years since Graduation" for each surgeon at time of entry



From Tables 3.5 a-c we note that a higher focus is associated with both fewer articles and fewer cites for the surgeons in our dataset but that result is only statistically significant for the early career surgeons for articles in all models. The coefficient for this variable is marginally statistically significant for cite-weighted articles for the Task Based Focus HHI. This implies that variation early in the career is associated with a higher quantity of articles but not once these articles are quality weighted. Lower clinical focus is not associated with more articles once the surgeon has accumulated a certain amount of experience. It is important to point out that the significance of the coefficients on the focus variable on the quantity of articles varies with the type of measure used. The level of significance goes down with the CABG/Valve based measure in which the categories are much less detailed. We consider the Task-Based Focus categories to be the most detailed and the CABG/Valve the least.

Tables 3.5 a-c. Herfindahl-Hirschmann Index Models

Table 3.5a. Fixed Effects Poisson Quasi Maximum Likelihood Model. Focus Measure - HHI of Task Based Focus Categories. P-values under coefficients. Robust standard errors clustered by surgeon

Dependent Variable:	Cite-Weighted Articles			Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Task Based Focus HHI	-0.385	-1.618*	0.102	-0.214	-1.800**	0.638
	(0.506)	(0.053)	(0.859)	(0.716)	(0.030)	(0.335)
Share Non-Cardiac Patients	-0.255	0.209	-0.322	-0.403	0.130	-0.316
	(0.684)	(0.896)	(0.592)	(0.286)	(0.864)	(0.455)
Total Number of Patients	0.00303***	0.00264***	0.00437***	0.00238***	0.0016***	0.00398***
	(0.007)	(0.004)	(0.000)	(0.003)	(0.005)	(0.001)
Cardiac Complexity (TBF)	0.295	-1.577*	1.003	-0.486	-2.232***	0.292
	(0.686)	(0.059)	(0.241)	(0.355)	(0.002)	(0.614)
Years Since Graduation	0.002	-0.018	0.029	0.007	0.011	0.027
	(0.941)	(0.638)	(0.333)	(0.676)	(0.504)	(0.254)
Number of Unique Surgeons	115	53	62	118	55	63
Number of Observations	952	397	555	964	406	558

*** p<0.01, ** p<0.05, * p<0.1

Table 3.5b. Fixed Effects Poisson Quasi Maximum Likelihood Model. Focus Measure - HHI of CABG/Valve Based Focus Categories. P-values under coefficients. Robust standard errors clustered by surgeon

Dependent Variable:	Cite-Weighted Articles			Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
CABG/Valve Based HHI	-0.266	-1.143	-0.14	-0.373	-1.197*	-0.098
	(0.750)	(0.289)	(0.884)	(0.549)	(0.073)	(0.915)
Share Non-Cardiac Patients	-0.213	0.146	-0.387	-0.430	0.060	-0.489
	(0.753)	(0.926)	(0.557)	(0.274)	(0.934)	(0.294)
Total Number of Patients	0.00309***	0.00252***	0.00439***	0.00244***	0.00147***	0.00390***
	(0.007)	(0.008)	(0.000)	(0.003)	(0.009)	(0.000)
Cardiac Complexity (TBF Based)	0.399	-1.311	0.863	-0.541	-1.870***	-0.131
	(0.614)	(0.178)	(0.340)	(0.263)	(0.009)	(0.829)
Years Since Graduation	0.001	-0.026	0.031	0.007	0.002	0.030
	(0.974)	(0.478)	(0.336)	(0.686)	(0.924)	(0.206)
Number of Unique Surgeons	115	53	62	118	55	63
Number of Observations	952	397	555	964	406	558

*** p<0.01, ** p<0.05, * p<0.1

Table 3.5c. Fixed Effects Poisson Quasi Maximum Likelihood Model. Focus Measure - HHI of CCS Based Focus Categories. P-values under coefficients. Robust standard errors clustered by surgeon

Dependent Variable:	Cite-Weighted Articles			Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
CCS Based Focus HHI	-0.213	-1.396	0.239	-0.344	-1.622**	0.346
	0.758	0.190	0.760	0.587	0.039	0.691
Share Non-Cardiac Patients	-0.222	0.412	-0.341	-0.373	0.266	-0.334
	0.712	0.792	0.602	0.306	0.738	0.484
Total Number of Patients	0.00298***	0.00256***	0.00430***	0.00241***	0.00157***	0.00394***
	0.008	0.010	0.000	0.003	0.004	0.001
Cardiac Complexity (CCS Based)	0.292	-1.214	0.989	-0.481	-2.180**	0.309
	0.730	0.135	0.386	0.415	0.014	0.664
Years Since Graduation	0.002	-0.025	0.031	0.006	0.004	0.027
	0.937	0.476	0.313	0.715	0.793	0.243
Number of Unique Surgeons	115	53	62	118	55	63
Number of Obs	952	397	555	964	406	558

*** p<0.01, ** p<0.05, * p<0.1

Another interesting result is that for the early career group, higher cardiac complexity, measured here as the average number of cardiac procedure categories per patient, is associated with a lower number of articles in all models and fewer cite-weighted articles in one of the models. The only variable that is associated very strongly with higher number of raw and cite-weighted articles is the number of cardiac patients. It is somewhat surprising that a higher number of patients and articles are positively correlated as both writing and operating are time-consuming activities. It implies that doing more surgeries is the only consistent predictor of increased innovative outcomes as measured by articles published which hints at a "learning to innovate by doing more routine work" phenomenon.

Our next set of results in Tables 3.6a-3.6c is from a sample of surgeons who do predominantly cardiac procedures. Excluded are all surgeons who over their years in our sample have on average more than 50% of their patients receiving only non-cardiac procedures such as

lung procedures, other thoracic or vascular procedures and so on. Since we only use each surgeon's cardiac procedures to construct the HHI variables in each model, we believe that the excluded surgeons are not necessarily comparable to the predominantly cardiac surgeons and this specification gives us a better understating of the effect of clinical focus among predominantly cardiac surgeons.

Our results are similar to those based on the whole sample. One difference is that a higher focus measured by the HHI is associated with a lower number of cite-weighed articles among early-career surgeons not only using the Task Based categories but also the CCS based categories and the results are highly significant. Also, a higher focus negatively impacts cite-weighted articles in the whole sample models, rather than just the early career samples, in both the Task Based Focus and the CABG/Valve models.

Table 3.6a: Fixed Effects Poisson Quasi Maximum Likelihood Model. Focus Measure is HHI of Task Based Focus Categories. Sample Restricted to surgeons with more than 50% of their patients receiving cardiac surgery. P-values under coefficients. Robust standard errors clustered by surgeon

Dependent Variable:	Cite-Weighted Articles			Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Task Based Focus HHI	-1.417*	-2.715***	-0.622	-1.063	-3.330***	0.112
	0.065	0.000	0.582	0.192	0.002	0.909
Share Non-Cardiac Patients	-0.984*	-0.666	-0.89	-0.820**	-0.523	-0.597
	0.083	0.721	0.124	0.038	0.6	0.226
Total Number of Patients	0.00246***	0.00280***	0.00337**	0.00221***	0.00172***	0.00367***
	0.005	0.002	0.013	0.004	0.007	0.007
Cardiac Complexity	-0.244	-1.699**	0.418	-0.625	-2.615***	0.266
	0.727	0.041	0.599	0.286	0.000	0.676
Years Since Graduation	-0.007	-0.022	0.013	0.001	0.014	0.016
	0.772	0.551	0.708	0.946	0.387	0.474
Number of Groups	107	47	60	109	48	61
Observations	881	348	533	888	352	536

*** p<0.01, ** p<0.05, * p<0.1

Table 3.6b: Fixed Effects Poisson Quasi Maximum Likelihood Model. Focus Measure - HHI of CABG/Valve Based Focus Categories. Sample Restricted to surgeons with more than 50% of their patients receiving cardiac surgery. P-values under coefficients. Robust standard errors clustered by surgeon

Dependent Variable:	Cite-Weighted Articles			Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
CABG/Valve Based HHI	-1.574**	-1.829	-1.527	-1.268*	-1.790*	-1.034
	0.0458	0.111	0.168	0.062	0.063	0.267
Share Non-Cardiac Patients	-1.161**	-0.716	-1.180*	-0.954**	-0.387	-0.874*
	0.048	0.714	0.061	0.011	0.703	0.091
Total Number of Patients	0.0026***	0.0026***	0.0036***	0.0023***	0.0014**	0.0038***
	0.006	0.006	0.007	0.003	0.013	0.002
Cardiac Complexity	-0.222	-1.275	0.162	-0.617	-2.004***	-0.12
	0.762	0.195	0.852	0.254	0.006	0.863
Years Since Graduation	-0.014	-0.035	0.012	-0.003	-0.001	0.018
	0.597	0.336	0.728	0.839	0.964	0.419
Observations	881	348	533	888	352	536
Number of Groups	107	47	60	109	48	61

*** p<0.01, ** p<0.05, * p<0.1

Table 3.6c: Fixed Effects Poisson Quasi Maximum Likelihood Model. Focus Measure - HHI of CCS Based Focus Categories. Sample Restricted to surgeons with more than 50% of their patients receiving cardiac surgery. P-values under coefficients. P-values under coefficients. Robust standard errors clustered by surgeon

Dependent Variable:	Cite-Weighted Articles			Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
CCS Based Herfindahl	-1.351	-2.942***	-0.577	-1.437*	-3.020**	-0.592
	0.127	0.009	0.618	0.061	0.019	0.517
Share Non-Cardiac Patients	-0.948	-0.488	-0.939	-0.796*	-0.257	-0.616
	0.113	0.792	0.200	0.066	0.806	0.328
Total Number of Patients	0.00236**	0.00266***	0.00332**	0.00217***	0.00156***	0.00357***
	0.013	0.005	0.012	0.006	0.004	0.008
Cardiac Complexity CCS	-0.412	-1.497*	0.175	-0.714	-2.561***	0.155
	0.606	0.074	0.879	0.248	0.007	0.837
Years Since Graduation	-0.009	-0.035	0.015	-0.002	0.002	0.016
	0.730	0.348	0.661	0.915	0.906	0.480
Observations	881	348	533	888	352	536
Number of Groups	107	47	60	109	48	61

*** p<0.01, ** p<0.05, * p<0.1

The other important difference is that we find that a higher share of non-cardiac patients is negatively associated with number of articles in all focus measure specifications and cite-weighted articles in the CABG/Valve focus measure specification. This implies that for those surgeons that are not focusing on non-cardiac areas having a higher portion of non-cardiac patients is associated with lower innovative output. This implies that specialization in cardiac surgery is good for innovative output if the surgeon is already specialized in cardiac surgery by having more than 50% of his patients already in cardiac surgery.

In the rest of our models we use a different measure of specialization – we include the share of each of the specific categories in the surgeon's procedure or patient (in CABG/valve) volume. The omitted category is Pacemaker/Defibrillator in the Task Based and the CCS categories and Cardiac without CABG or Valve in the CABG/Valve categories.

We find, in Table 3.7a that an increase in the share of all categories at the expense of the Pacemaker/Defibrillator category is associated with an increased number of cite-weighted articles for the late-career surgeons and the whole sample of surgeons. However, the result is not significant for the Heart Valve category for late-career surgeons and Heart Valve and CABG categories for the whole sample. This result is probably due to the fact that the procedures in the Pacemaker-Defibrillator category are the least complex ones and as a result do not allow for the publication of many articles.

Table 3.7a: Fixed Effects Poisson Quasi Maximum Likelihood Model. Task Based Focus Categories. P-values under coefficients. Robust standard errors clustered by surgeon

Dependent Variable:	Cite-Weighted Articles			Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Share CABG Procedures	1.430	-1.212	3.467***	0.589	0.083	1.817*
	0.103	0.306	0.000	0.369	0.944	0.086
Share Valve Procedures	0.627	-1.522	2.325	1.084	-0.162	1.940
	0.654	0.560	0.190	0.347	0.924	0.178
Share HTV Procedures	2.294*	0.241	4.724**	0.829	1.385	-0.035
	0.097	0.895	0.031	0.453	0.194	0.989
Share Heart Muscle Procedures	1.840**	0.319	3.704***	1.091*	0.114	2.210**
	0.036	0.847	0.000	0.091	0.918	0.022
Share Heart Assist Transplant Proc	3.189*	-2.907	6.348***	1.618	-2.396**	2.467
	0.064	0.301	0.002	0.103	0.031	0.243
Share Non-cardiac Patients	-0.540	-0.495	-0.587*	-0.574	-0.137	-0.468
	0.300	0.787	0.094	0.108	0.889	0.199
Total Patients	0.0033***	0.0027***	0.00491***	0.00250***	0.00137*	0.00427***
	0.001	0.010	0.000	0.002	0.056	0.001
Cardiac Complexity	0.438	-0.253	0.616	-0.507	-0.823	0.253
	0.622	0.824	0.635	0.440	0.257	0.815
Years Since Graduation	0.015	-0.046	0.044	0.003	-0.021	0.028
	0.565	0.268	0.141	0.894	0.425	0.245
Number of Groups	115	53	62	118	55	63
Number of Obs	952	397	555	964	406	558

*** p<0.01, ** p<0.05, * p<0.1

It is surprising that trading off any of the categories against the easy Pacemaker/Defibrillator category does not result in higher innovative output for early-career surgeons. In fact one of the most complex categories Heart Assists and Transplants is associated with a lower number of articles for early career surgeons and the result is statistically significant. This implies that a larger variety of simpler procedures is correlated with a higher innovative output for early career surgeons.

Table 3.7b: Fixed Effects Poisson Quasi Maximum Likelihood Model. CABG/Valve Based Focus Categories. P-values under coefficients. Robust standard errors clustered by surgeon

Dependent Variable:	Cite-Weighted Articles			Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Share CABG Patients	-0.079	-1.848	0.346	-0.235	-0.081	-0.039
	0.854	0.128	0.241	0.542	0.917	0.948
Share Valve Patients	-1.445*	-3.152	-1.204	-0.181	0.189	-0.188
	0.091	0.210	0.297	0.811	0.870	0.870
Share CABG and Valve Patients	-0.307	3.646	0.287	0.649	-1.348	2.539
	0.900	0.636	0.875	0.720	0.583	0.216
Share Noncardiac Patients	-0.383	-1.315	-0.280	-0.520	-0.032	-0.506
	0.510	0.567	0.519	0.173	0.973	0.247
Total Patients	0.00289***	0.002	0.00428***	0.00233***	0.00148*	0.00407***
	0.008	0.112	0.000	0.005	0.051	0.001
Cardiac Complexity	0.828	-1.161	1.095	-0.535	-0.750	-0.634
	0.373	0.600	0.183	0.327	0.434	0.340
Years Since Graduation	0.017	-0.021	0.048	0.004	0.001	0.032
	0.534	0.489	0.106	0.831	0.973	0.207
Number of Groups	115	53	62	118	55	63
Number of Obs	952	397	555	964	406	558

*** p<0.01, ** p<0.05, * p<0.1

In the CABG/Valve Category results, we see that a higher share of Valve Patients at the expense of Cardiac without CABG or Valve is associated with a significantly lower number of cite-weighted articles. In fact even a higher share of CABG is associated with a lower number of cite-weighted articles, even though the result is insignificant implying that it is the rarer and potentially more complex procedures on patients that do not receive CABG and/or Valve that increase innovative output as measured by cite-weighted articles

In the CCS category, late-career surgeon results are consistent with those from the Task Based Focus category model – increases in the shares of CABG and Other Cardiac OR

procedures (but not Valve) at the expense of Pacemaker/Defibrillator are associated with a higher number of cite-weighted articles.

Table 3.7c: Fixed Effects Poisson Quasi Maximum Likelihood Model. CCS Based Focus Categories. P-values under coefficients. Robust standard errors clustered by surgeon

Dependent Variable:	Cite-Weighted Articles			Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Share CABG Procedures	0.793	-1.971	2.669***	0.463	0.303	1.432
	0.332	0.134	0.000	0.460	0.769	0.163
Share Valve Procedures	-0.268	-1.198	1.130	0.991	0.707	1.970
	0.844	0.628	0.526	0.403	0.652	0.231
Share Other OR Procedures	0.915	-1.356	2.617***	1.018*	-0.134	2.087**
	0.255	0.560	0.001	0.074	0.902	0.021
Share Noncardiac CCS Patients	-0.305	0.139	-0.357	-0.565*	0.285	-0.603*
	0.544	0.937	0.337	0.089	0.768	0.066
Total Patients	0.00307***	0.00245**	0.00495***	0.00244***	0.00132**	0.00429***
	0.008	0.013	0.000	0.004	0.029	0.002
Cardiac Complexity	0.823	-0.763	1.452*	-0.455	-1.253	-0.073
	0.273	0.455	0.067	0.370	0.108	0.912
Years Since Graduation	0.008	-0.029	0.043	-0.003	-0.004	0.021
	0.763	0.343	0.169	0.881	0.859	0.425
Number of Groups	115	53	62	118	55	63
Number of Observations	952	397	555	964	406	558

*** p<0.01, ** p<0.05, * p<0.1

The last two sets of models in Tables 3.8a-c and 3.9a-c have as outcome variables raw and cite-weighted articles in specific fields – CABG and Heart Valve respectively. You will note that the number of observations is fewer and that is because our models drop surgeons who do not have any articles or any cites over their time in our dataset in the specific CABG and Heart Valve fields. We are specifically interested in these models to understand how work in related fields influences innovative output in a specific category.

3.8 Models Based on Share of Total Procedures in Category. CABG Articles Only

Table 3.8a: Fixed Effects Poisson Quasi Maximum Likelihood Model. Task Based Focus Categories. CABG Articles Only. P-values under coefficients. Robust standard errors clustered by surgeon.

Dependent Variable:	Cite-Weighted CABG Articles			CABG Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Share CABG Procedures	-0.554	-2.612***	3.031**	0.004	0.292	0.985
	0.790	0.000	0.015	0.998	0.904	0.615
Share Valve Procedures	4.858	-3.584	11.53***	1.111	-5.881	3.875
	0.110	0.448	0.000	0.644	0.185	0.161
Share HTV Procedures	4.714	-5.598	16.10**	-3.856	-0.830	-6.514
	0.327	0.259	0.011	0.204	0.826	0.175
Share Heart Muscle Procedures	2.562	-2.023	7.596***	1.754	0.459	2.983
	0.415	0.495	0.008	0.366	0.879	0.224
Share Heart Assist Procedures	2.890	-9.230	10.62*	1.268	-3.571	1.140
	0.523	0.177	0.070	0.706	0.474	0.803
Share Non-cardiac Patients	-0.503	2.970	-1.345	-1.195	-4.170	-0.114
	0.722	0.211	0.341	0.547	0.244	0.952
Total Patients	0.00366**	0.00806***	0.00533***	0.00316**	0.001	0.00397*
	0.034	0.001	0.008	0.040	0.602	0.060
Cardiac Complexity	1.070	2.500	1.027	2.575	2.111	4.272*
	0.595	0.416	0.637	0.156	0.420	0.078
Years Since Graduation	0.011	-0.019	-0.004	0.021	0.079	-0.019
	0.839	0.841	0.947	0.592	0.158	0.731
Number of Groups	70	29	41	74	32	42
Number of Observations	605	228	377	638	256	382

*** p<0.01, ** p<0.05, * p<0.1

We find that in the Task Based and CCS Based Category models, an increase in the share of any category at the expense of Pacemaker/Defibrillator increases the number of cite-weighted CABG articles (but not raw articles) for the late-career surgeons. However, for early-career surgeons, a higher number of CABG procedures at the expense of the respective omitted category in each table is associated with a lower number of cite-weighted CABG articles. This result is surprising. But given the fact that CABG surgeries constitute the vast majority of cardiac

surgeries for most surgeons it may imply that in diversifying away from CABG is what leads to more article in CABG, potentially hinting at cross-pollination between CABG and other procedures.

Table 3.8b: Fixed Effects Poisson Quasi Maximum Likelihood Model. CABG/Valve Based Focus Categories. CABG Articles Only. P-values under coefficients. Robust standard errors clustered by surgeon.

Dependent Variable:	Cite-Weighted CABG Articles			CABG Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Share CABG Patients	-0.738	-2.011*	0.237	-0.584	0.336	-0.691
	0.605	0.052	0.872	0.648	0.844	0.671
Share Valve Patients	4.454**	2.152	6.665***	0.777	-1.173	1.921
	0.026	0.574	0.002	0.657	0.697	0.368
Share CABG and Valve Patients	-2.007	-6.460	-1.379	-0.441	-7.771	2.620
	0.469	0.579	0.600	0.894	0.134	0.405
Share Non-cardiac Patients	0.492	3.520**	1.157	-1.105	-3.337	-0.188
	0.734	0.034	0.478	0.578	0.326	0.931
Total Patients	0.00356**	0.00961***	0.004	0.00298**	0.002	0.003
	0.037	0.000	0.123	0.040	0.409	0.121
Cardiac Complexity	3.128*	1.051	4.668***	2.439	2.105	2.873*
	0.067	0.761	0.005	0.104	0.479	0.086
Years Since Graduation	-0.008	-0.022	-0.033	0.019	0.072	-0.018
	0.900	0.817	0.679	0.644	0.194	0.765
Number of Groups	70	29	41	74	32	42
Number of Observations	605	228	377	638	256	382

*** p<0.01, ** p<0.05, * p<0.1

Also surprisingly, an increase in the Valve category in all models is associated with a higher number of CABG articles. That a result is not statistically significant in the Task Based Focus models and is really driven by late-career surgeons. This implies that there is cross-pollination from the Heart Valve to the CABG categories meaning that doing more Heart Valve procedures may increase surgeons' insight in CABG surgeries and subsequently help them publish articles in that field. But it is also important to remember that our omitted variable is

Pacemaker/Defibrillator so this result means that if innovative outcomes in CABG are the desired outcome, then surgeons should substitute Heart Valve procedures rather than CABG procedures for Pacemaker/Defibrillator procedures.

Table 3.8c: Fixed Effects Poisson Quasi Maximum Likelihood Model. CCS Based Focus Categories. CABG Articles Only. P-values under coefficients. Robust standard errors clustered by surgeon.

Dependent Variable:	Cite-Weighted CABG Articles			CABG Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Share CABG Procedures (CCS based)	-0.289	-3.192***	3.797**	0.009	-0.118	1.202
	0.898	0.000	0.011	0.996	0.962	0.551
Share Valve Procedures (CCS Based)	5.059*	-1.592	11.21***	1.366	-5.092	4.609*
	0.085	0.654	0.000	0.569	0.196	0.089
Share Other OR Procedures (CCS Based)	3.010	-0.908	7.971**	2.116	1.065	3.471
	0.386	0.836	0.021	0.330	0.754	0.222
Share Noncardiac (CCS Based)	-0.218	2.946	-0.259	-1.456	-4.427	-0.397
	0.863	0.206	0.852	0.459	0.258	0.842
Total Patients	0.00388**	0.00846***	0.00622***	0.00307**	0.001	0.00401*
	0.036	0.000	0.001	0.042	0.543	0.072
Cardiac Complexity	1.441	0.353	2.531	1.947	1.441	3.108*
	0.419	0.894	0.172	0.181	0.537	0.091
Years Since Graduation	0.006	0.003	-0.005	0.018	0.087	-0.018
	0.915	0.968	0.942	0.649	0.109	0.750
Number of Groups	70	29	41	74	32	42
Number of Observations	605	228	377	638	256	382

*** p<0.01, ** p<0.05, * p<0.1

The results from the models that have Heart Valve articles as a dependent variable in Tables 3.9a-c are less surprising. We find that increasing the share of any cardiac procedure at the expense of Pacemaker/Defibrillator increases cite-weighted Heart Valve articles. The results are statistically significant for the whole sample and late-career surgeons sample of the Heart For early-career surgeons, it is increasing experience in the specific category i.e. Heart Valve that is related to a higher number of cite-weighted Heart Valve articles in all models. This may

imply that insights gained doing other kind of cardiac surgery are less easily transferrable to Heart Valve related innovations.

3.9. Models Based on Share of Total Procedures in Category. Valve Articles Only

Table 3.9a: Fixed Effects Poisson Quasi Maximum Likelihood Model. Task Based Focus Categories. Valve Articles Only. P-values under coefficients. Robust standard errors clustered by surgeon.

Dependent Variable:	Cite-Weighted Valve Articles			Valve Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Share CABG Procedures	6.316***	4.062	9.033**	2.363	1.322	1.946
	0.002	0.136	0.019	0.221	0.652	0.553
Share Heart Valve Procedures	4.634***	11.29***	5.519*	1.099	2.917	0.654
	0.003	0.000	0.063	0.611	0.290	0.852
Share Heart Vessel Procedures	4.945*	5.785*	8.869**	-4.559	-0.874	-5.913
	0.065	0.069	0.048	0.196	0.816	0.353
Share Heart Muscle Procedures	8.006***	1.426	10.88***	1.156	-3.771	2.270
	0.000	0.694	0.001	0.604	0.231	0.547
Share Heart Assist Procedures	0.735	4.136	4.762	-4.405	-4.545	-4.682
	0.890	0.309	0.685	0.330	0.242	0.630
Share Non-cardiac Patients	-8.256***	-6.801*	-8.160**	-1.599	2.628	-3.083*
	0.001	0.066	0.018	0.333	0.392	0.082
Total Patients	0.00563**	0.00597**	0.006	0.002	0.00388*	0.001
	0.014	0.032	0.101	0.198	0.070	0.712
Cardiac Complexity	-0.035	-1.425	-0.445	1.250	-0.451	1.115
	0.979	0.347	0.800	0.284	0.811	0.487
Years Since Graduation	0.154***	0.123*	0.169***	0.116***	0.106	0.112**
	0.000	0.076	0.005	0.003	0.120	0.029
Number of Groups	52	23	29	55	23	32
Number of Observations	439	168	271	461	168	293

*** p<0.01, ** p<0.05, * p<0.1

It is also important to note that an increase in the number of Non-Cardiac procedures is related to a lower number of Heart Valve articles. Also our career-age variable is significant and positive in all of the models meaning that more experience is the only variable that is consistently related to a higher number of Heart Valve articles.

Table 3.9b: Fixed Effects Poisson Quasi Maximum Likelihood Model. CABG/Valve Based Focus Categories. Valve Articles Only. P-values under coefficients. Robust standard errors clustered by surgeon.

Dependent Variable:	Cite-Weighted Valve Articles			Valve Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All Surgeons	Early Career Surgeons	Late Career Surgeons
Share CABG Patients	-0.191	1.668	-0.690	0.688	2.634*	-0.541
	0.927	0.415	0.809	0.508	0.057	0.718
Share Valve Patients	-1.088	5.779***	-2.810	-0.172	1.512	-1.087
	0.571	0.002	0.225	0.914	0.544	0.527
Share CABG and Valve Patients	4.466	11.89**	1.350	7.511***	14.01***	4.798
	0.159	0.036	0.686	0.006	0.002	0.107
Share Non-cardiac Patients	-8.057***	-6.152**	-8.556**	-1.884	0.193	-2.955
	0.003	0.049	0.018	0.261	0.944	0.136
Total Patients	0.00436**	0.005	0.004	0.002	0.003	0.000
	0.045	0.107	0.181	0.191	0.198	0.977
Cardiac Complexity	-1.934	-3.968*	-1.360	-2.162**	-3.583**	-1.986*
	0.139	0.068	0.376	0.021	0.017	0.095
Years Since Graduation	0.151***	0.126**	0.170***	0.128***	0.140**	0.114**
	0.000	0.041	0.003	0.000	0.019	0.022
Number of Groups	52	23	29	55	23	32
Number of Observations	439	168	271	461	168	293

*** p<0.01, ** p<0.05, * p<0.1

Table 3.9c: Fixed Effects Poisson Quasi Maximum Likelihood Model. CCS Based Focus Categories. Valve Articles Only. P-values under coefficients. Robust standard errors clustered by surgeon.

Dependent Variable:	Cite-Weighted Valve Articles			Valve Articles		
	All Surgeons	Early Career Surgeons	Late Career Surgeons	All SurgeonsAh	Early Career Surgeons	Late Career Surgeons
Share CABG Procedures	5.648***	3.934	8.445**	2.560	2.924*	1.684
	0.005	0.107	0.031	0.193	0.057	0.637
Share Heart Valve Procedures	4.845***	10.63***	5.400*	2.441	4.787*	1.526
	0.009	0.000	0.086	0.297	0.056	0.673
Share Other OR Cardiac Procedures	7.746***	2.000	10.24***	2.205	-0.896	2.869
	0.000	0.590	0.004	0.358	0.632	0.484
Share Non-cardiac Procedures (CCS Based)	-8.537***	-6.546*	-8.231**	-1.993	1.487	-2.972
	0.001	0.055	0.033	0.245	0.610	0.144
Total Patients	0.00535**	0.00599**	0.005	0.00269*	0.00462***	0.001
	0.014	0.028	0.113	0.052	0.008	0.737
Cardiac Complexity	-0.828	-1.533	-0.617	-0.282	-0.476	-0.763
	0.484	0.251	0.664	0.779	0.668	0.511
Years Since Graduation	0.153***	0.127**	0.162***	0.120***	0.123**	0.103**
	0.000	0.043	0.007	0.001	0.018	0.044
Number of Groups	52	23	29	55	23	32
Number of Observations	439	168	271	461	168	293

*** p<0.01, ** p<0.05, * p<0.1

3.6 Discussion

In this paper we tried to answer the question: "Does the composition of routine work, specifically its diversity, influence innovative outcomes?" Our results indicate that innovation by early-career workers may benefit from a diverse set of routine tasks. We also find that innovation in certain tasks (Heart Valve) may benefit more from experience in the focal task while innovation in other tasks (CABG) may benefit from related experience.

We tested our results on a sample of cardiac surgeons whose work is based on constant fast paced decision making in life and death situations. Even the most routine procedures in this setting require alertness and continuous problem solving. Arguably, each patient is unique and

may require modifications of the procedure on the spot.[12] As such, our results may not necessarily be generalizable to settings where tasks are more monotonous and standardized and alterations are not needed– for example, entering a form into the computer. However, we believe that the results apply to situations where the inputs are not identical, where problem solving and decision making are necessary and innovation is encouraged.

One of the limitations of our work is that we measure focus and diversity of work as a yearly variable for each surgeon. It is possible (not likely) that a surgeon may perform only CABGs in the spring, only Heart Valve procedures in the summer and so on. A better measure of how insight from one task could be transferred to another or how multiple identical procedures in a row are better for gaining a new insight would be to track how tasks alternate, i.e. whether a Heart Valve procedure is followed by a CABG procedure and so on.

We believe that this work is important as a first attempt to understand the importance of work specialization and diversity on innovative outcomes at the individual level. Our various results are difficult to interpret and do not always point to one answer but instead show that many different variables such as procedure complexity can be important. We hope that future work will elucidate better the mechanisms by which work diversity or focus may influence innovation. We also hope that future research will explore other settings in which the results may be corroborated or modified.

[12] Note however, that some may argue that not all patients are unique and that procedures and patient care can also be standardized.

3.7 Bibliography

Adams, J. and Z. Griliches (1996), "Measuring Science: An Exploration," *Proceedings of the National Academy of Sciences of the United States Of America*, vol. 93, no. 23, pp. 12664-12670

Adler, P.S. and K.B. Clark (1991), "Behind the Learning-Curve – a Sketch of the Learning Process," *Management Science*, vol. 37, no. 3, pp. 267-281

Ancona, D and D. Caldwell (1992), "Demography and Design: Predictors of New Product Team Performance." *Organization Science*, vol.3, pp. 321-341

Argote, L. and D. Epple (1990), "Learning Curves in Manufacturing," *Science*, vol. 247, no. 4945, pp. 920-924

Argote, L., Beckman, S. and D. Epple (1990), "The Persistence and Transfer of Learning in Industrial Settings," *Management Science, vol.* 36, pp.140-154

Arrow, K.J. (1962), "The Economic Implications of Learning by Doing," *Review of Economic Studies*, vol.29, no.80, pp. 155-173

Astebro T. and P. Thompson (2011), "Entrepreneurs: Jacks of all Trades or Hobo?", *Research Policy*, vol. 40, no.5, pp. 637-649

Azoulay, P. (2002), "Do pharmaceutical sales respond to scientific evidence?", *Journal of Economics & Management Strategy*, vol. 11, no. 4, pp. 551-594

Berndt, E.R., Cockburn, I.M. and K. A. Grepin (2006), "The Impact of Incremental Innovation on Biopharmaceuticals: Drug Utilization in Original and Supplemental Indications", P*harmacoEconomics*, vol. 26, no. 2, pp.69-86

Ball, C.G., Sutherland, F., Kirkpatrick, A.W. , Dixon, E. ,MacLean, A.R., Mack, L.A., Feliciano, D.V., Rajani, R.R., Karmy-Jones, R., Buie, W.D. , Temple, W.J., Rozycki, G.S. and A. Simeone (2010), "Dramatic innovations in modern surgical subspecialties," *Canadian Journal of Surgery*, vol.53, no.5, pp. 335-341

Bettenhausen, K.L. (1991), "Five Years of Groups Research: What We Have Learned and What Needs to Be Addressed," *Journal of Management* vol.17, no.2, pp. 345–381.

Boh, W. F., Slaughter, S. A. and J. A. Espinosa (2007), "Learning from Experience in Software

Development: a Multilevel Analysis," *Management Science, vol.* 53, no.8, pp. 1315-1331

Brush, T. and A. Karnani (1996), "Impact of Plant Size and Focus on Productivity: An Empirical Study," *Management Science*, vol.42, no.7, pp.1065-1081

Bunderson, J.S. and K.M. Suthcliffe (2002), “Comparing alternative conceptualizations of functional diversity in management teams: Process and performance effects,” *Academy of Management Journal*, vol. 45, no.5, pp. 875-893

Chatterji, A.K., Fabrizio, K.R., Mitchell, W. and K.A. Schulman (2008), “Physician-Industry Cooperation in the Medical Device Industry,” *Health Affairs*, vol. 27, no.6, pp.1532-1643

Cockburn, I., Bailit, H., Berndt, E. and S. Finkelstein (1999), “When Antihistamines Go To Work,” *Business and Health*, vol. 17, pp. 3, pp. 49-5

Cockburn, I. and R. Henderson (1998), “Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery,” *Journal of Industrial Economics*, vol. 46, no.2, pp. 157-182

Cockburn, I., and R. Henderson (2001), "Scale and Scope in Drug Development: Unpacking the Advantages of Size in Pharmaceutical Research," *Journal of Health Economics*, vol.20, no.6, pp.1033-1057

Christensen, B.T. And C. D. Schunn (2007), “The relationship of analogical distance to analogical function and preinventive structure: The case of engineering design,” *Memory & Cognition Vol. 35, no.1,pp. 29-38*

Diwas, K.C. & B.R. Staats (Forthcoming), “Accumulating a Portfolio of Experience: The Effect of Focal and Related Experience on Surgeon Performance,” *Manufacturing & Service Operations Management*

Edmondson, A.C., Winslow, A.B., Bohmer, R.M.J. and G. Pisano (2003), “Learning How and Learning What: Effects of Tacit and Codified Knowledge on Performance Improvement Following Technology Adoption,” *Decision Sciences*, vol. 34, no. 2, pp. 197-223

Gauderer, M.W.L. (2009), “Creativity and the Surgeon,” *Journal of Pediatric Surgery*, vol. 44, pp.13–20

Hall, B.L., Hsiao, E.Y., Majercik, S., Hirbe, M. and B.H. Hamilton (2009, “The Impact of Surgeon Specialization on Patient Mortality: Examination of a Continuous Herfindahl-Hirschman Index,*” Annals of Surgery*, vol. 249, no. 5, pp.708-716

Hong, L. and S.E. Page (2004), “Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers,” Proceedings of the National Academy of Sciences, pp.16385-16389

Huckman, R.S. and G.P. Pisano (2006), “The Firm Specificity of Individual Performance: Evidence from Cardiac Surgery,” *Management Science*, vol. 52, no.4, pp.473-488

Huckman R.S. and D. Zinner (2007), “Does Focus Improve Operational Performance? Lessons from the Management of Clinical Trials,” *Strategic Management Journal*, vol.29, no.2, pp.173-193

Hyysalo, S.(2006), “The Role of Learning-by-Using in the Design of Health Care Technologies: a Case Study,” *Information Society*, vol. 22, no. 2, pp.599-626

Johnson, M.K. and L. Hasher (1987),”Human Learning and Memory,” *Annual Review of Psychology*, vol. 38, pp.631-668

Kalogerakis K., Luethje, C. and C. Herstatt (2010), “Developing Innovations Based on Analogies: Experience from Design and Engineering Consultants,” Journal of Product and Innovation Management, vol. 27, pp. 418-436

Lapré, M. A. and I. M. Nembhard (2010), "Inside the Organizational Learning Curve: Understanding the Organizational Learning Process," *Foundations and Trends in Technology, Information and Operations Management*, vol.4, no.1, pp.1-103

Levitt, B. and J.G. March (1988), “Organizational Learning,” *Annual Review of Sociology*, vol.14, pp. 319-340

Lindsey, D. (1989), “Using Citation Counts as a Measure of Quality in Science - Measuring What’s Measurable rather Than What’s Valid,” *Scientometrics*, vol. 15, no. 3-4, pp.189-203

March, J.G. (1991), “Exploration and Exploitation in Organizational Learning,” *Organization Science*, vol.2, no.1, pp.71-87

Mehta, H. (2009), “On Innovations in Plastic Surgery,” *Journal of Plastic, Reconstructive and Aesthetic Surgery*, vol. 62, pp.437-441

Merton, R.K. (1957), “Priorities in Scientific Discovery – a Chapter in the Sociology of Science,” *American Sociological Review*, vol.22, pp. 635-659

Narayanan, S., Balasubramanian, S. and J. M. Swaminathan (2009), "A Matter of Balance: Specialization, Task Variety, and Individual Learning in a Software Maintenance Environment" *Management Science*, vol.55, no.11, pp.1861-1876

Newell A & Simon H. A. (1972), Human problem solving, Englewood Cliffs, NJ: Prentice-Hall

Park, M.H.J., Lim, J.W. and P.H. Birnbaum-More (2009), “The Effect of Multiknowledge Individuals on Performance in Cross-Functional New Product Development Teams,” *Journal of Product Innovation Management*, vol. 26, no.1 pp. 86-96

Pelled, L.H., Eisenhardt, K.M., and K.R. Xin (1999), “Exploring the Black Box: An Analysis of Work Group Diversity, Conflict, and Performance,” *Administrative Science Quarterly*, vol. 44, no. 1, pp. 1–28

Rico, M., Molleman, E., Sanchez-Manzanares, M. and G.S. Van der Vegt (2007), “The effects of diversity faultlines and team task autonomy on decision quality and social integration,” *Journal of Management*, vol. 33, no. 1, pp. 111-132

Riskin, D.J., T. Longaker, M., Gertner, M. and T.M. Krummel (2006), “Innovation in Surgery – a Historical Perspective,” *Annals of Surgery*, vol. 244, no.5, pp.686-693

Rosenberg, N. (1979), “Technological Interdependence in the American Economy,” *Technology and Culture*, vol. 20, no. 1, pp.25-50

Schilling, M. A., Vidal,P., Ployhart R. E. and A. Marangoni (2003), "Learning by Doing Something Else: Variation, Relatedness, and the Learning Curve." *Management Science, vol.* 49, no.1, pp.39-56

Schroeder, S.A., Zones, J.S. and J. Showstack (1989), “Academic Medicine as a Public Trust,” *JAMA*, vol. 262, no.6, pp.803-812

Simon, H. A. (1990), “Invariants of human behavior,” *Annual Review of Psychology*, vol. 41, pp.1-20

Staats, B. R. & F.Gino (Forthcoming), “Specialization and Variety in Repetitive Tasks: Evidence from a Japanese Bank,” *Management Science*

Starr, J.P. (2010), “Tetralogy of Fallot: Yesterday and Today,” *World Journal of Surgery*, vol. 34, pp. 658-668

Stephan, P.E. (1996), “The Economics of Science,” *Journal of Economic Literature*, vol. 34, no.3, pp. 1199-1235

Tucker, A.L., Edmondson, A.C. and S. Spear (2002), “When Problem Solving Prevents Organizational Learning,” *Journal of Organizational Change Management*, vol.15, no.2, pp.122-137

Von Hippel, E. and M. J. Tyre (1995), “How Learning by Doing is Done – Problem Identification in Novel Process Equipment,” *Research Policy*, vol.24, no.1, pp.1-12

Vokurka, R. J., and R. A. Davis (2000), "Focused Factories: Empirical Study of Structural and Performance Differences," *Production and Inventory Management Journal*, vol. 41, no. 1, pp. 44-55

Williams, K.Y. and C.A. O'Reilly, III. (1998), "Demography and Diversity in Organizations: A Review of the 40 Years of Research," *Research in Organizational Behavior* vol. 20, no.20, pp. 77–140.

Wright, T.P. (1936), "Factors Affecting the Cost of Airplanes," *Journal of Aeronautical Sciences*, vol.3, no.4, pp.122–128

3.8 Appendixes

Appendix 3.A International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) procedure codes used in constructing the categories of the Task Based Focus measure

1. **Coronary Artery Bypass Grafting (CABG) Procedures**
   361, 3610, 3611, 3612, 3613, 3614, 3615, 3616, 3617, 3619

2. **Valve procedures**
   350, 3500, 3501, 3502, 3503, 3504, 351, 3510, 3511, 3512, 3513, 3514, 352, 3520, 3521, 3522, 3523, 3524, 3525, 3526, 3527, 3528, 3599

3. **Heart Muscle Procedures**
   353, 3531, 3532, 3533, 3534, 3535, 3539, 354, 3541, 3542, 355, 3550, 3551, 3552, 3553, 3554, 356, 3560, 3561, 3562, 3563, 357, 3570, 3571, 3572, 3573, 358, 3581, 3582, 3583, 3584, 359, 3591, 3592, 3593, 3594, 3595, 3598, 371, 3710, 3711, 3712, 3724, 3725, 373, 3731, 3732, 3733, 3735, 3749, 374

4. **Heart Vessel Procedures other than CABG**
   3603, 362, 363, 3631, 3632, 3639, 369, 3691, 3699, 3834, 3835, 3844, 3845

5. **Heart Assist and Heart Transplant Procedures**
   3741, 3752, 3753, 3754, 376, 3761, 3762, 3763, 3764, 3765, 3766, 3767, 3751, 336

6. **Pacemaker and Defibrillator Procedures**
   377, 3770, 3771, 3772, 3773, 3774, 3775, 3776, 3777, 3778, 3779, 378, 3780, 3781, 3782, 3783, 3784, 3785, 3786, 3787, 3789, 3794, 3795, 3796, 3797, 3798, 3799, 0050, 0051, 0052, 0053, 0054

Appendix 3.B1 Description of CCS level 2 Categories. Those Included in our focus measure are in bold. Category 16.1 is also included in our measure in category 7.7 and comes from section 16. Miscellaneous diagnostic and therapeutic procedures

Section 7. Operations on the cardiovascular system

CCS LVL 2	CCS LVL 2 LABEL
7.1	**Heart valve procedures [43.]**
7.2	**Coronary artery bypass graft (CABG) [44.]**
7.3	Percutaneous transluminal coronary angioplasty (PTCA) [45.]
7.4	Coronary thrombolysis [46.]
7.5	Diagnostic cardiac catheterization; coronary arteriography [47.]
7.6	**Insertion; revision; replacement; removal of cardiac pacemaker or cardioverter/defibr**
7.7	**Other OR heart procedures [49.]**
7.8	Extracorporeal circulation auxiliary to open heart procedures [50.]
7.9	Endarterectomy; vessel of head and neck [51.]
7.10	**Aortic resection; replacement or anastomosis [52.]**
7.11	Varicose vein stripping; lower limb [53.]
7.12	Other vascular catheterization; not heart [54.]
7.13	Peripheral vascular bypass [55.]
7.14	Other vascular bypass and shunt; not heart [56.]
7.15	Creation; revision and removal of arteriovenous fistula or vessel-to-vessel cannula
7.16	Hemodialysis [58.]
7.17	Other OR procedures on vessels of head and neck [59.]
7.18	Embolectomy and endarterectomy of lower limbs [60.]
7.19	Other OR procedures on vessels other than head and neck [61.]
7.20	Other diagnostic cardiovascular procedures [62.]
7.21	Other non-OR therapeutic cardiovascular procedures [63.]

Section 16: Miscellaneous diagnostic and therapeutic procedures

16.1	Other organ transplantation [176.]

Appendix 3.B2 International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) procedure codes used in constructing the categories and focus measure based on the Clinical Classification Software (CCS) system

7.1 Heart valve procedures 3500, 3501, 3502, 3503, 3504, 3505, 3506, 3507, 3508, 3509, 3510, 3511, 3512, 3513, 3514, 3520, 3521, 3522, 3523, 3524, 3525, 3526, 3527, 3528, 3596, 3597, 3599

7.2 Coronary artery bypass graft (CABG) - includes 7.10 Aortic resection; replacement or anastomosis marked in bold
3611, 3612, 3613, 3614, 3610, 3615, 3616, 3617, 3619, 362, 3620, 363, 3630, 3631, 3632, 3633, 3634, 3639, **3834, 3844, 3864, 3971, 3973, 3978**

7.6 Insertion; revision; replacement; removal of cardiac pacemaker or cardioverter/defibr
0050, 0051, 0052, 0053, 0054, 0056, 0057, 3770, 3771, 3772, 3773, 3774, 3775, 3776, 3777, 3778, 3779, 3826, 3780, 3781, 3782, 3783, 3785, 3786, 3787, 3789, 1751, 1752, 3794, 3795, 3796, 3797, 3798, 3531

7.7 Other OR heart procedures – also includes 16.1 Other organ transplantation , marked in bold 3532, 3533, 3534, 3535, 3539, 3541, 3542, 3550, 3551, 3552, 3553, 3554, 3555, 3560, 3561, 3562, 3563, 3570, 3571, 3572, 3573, 3581, 3582, 3583, 3584, 3591, 3592, 3593, 3594, 3595, 3598, 3600, 3603, 3609, 3691, 3699, 3710, 3711, 3712, 3731, 3732, 3733, 3734, 3735, 3736, 3737, 374, 3740, 3741, 3749, 3752, 3753, 3754, 3755, 3760, 3761, 3762, 3763, 3764, 3765, 3766, 3767, 3768, 3790, 3791, 3799, **375 , 3750, 3751, 336 , 3360**